

Lessons from the History and Philosophy of Science regarding the Research Assessment Exercise

DONALD GILLIES

1. Introduction

The Research Assessment Exercise (henceforth abbreviated to RAE) was introduced in 1986 by Thatcher, and was continued by Blair. So it has now been running for 21 years. During this time, the rules governing the RAE have changed considerably, and the interval between successive RAEs has also varied. These changes are not of great importance as far as the argument of this paper is concerned. We will concentrate on the main features of the RAE which can be summarised as follows.

At intervals of a few years, RAEs are carried out in all the universities of the UK. The first step is to appoint a committee of assessors in each subject. These assessors are usually academics working in the field in question in the UK. Next most members of each department in a subject have to select a set of pieces of their research. The department then submits all these pieces of research produced by its members to the assessment committee. The members of the committee study this research output, and, on its basis, grade the department on a scale running from very good downwards. The departments which score well on the RAE are provided with research funds. Those which don't score so well are less fortunate. They are provided with much smaller funds for research, and the members of such departments have to spend more time on teaching. Recently there have even been moves in some universities to close altogether departments which perform badly on the RAE.

Such then in rough outline is the procedure followed in the RAE. It should be pointed out that the RAE is a costly operation, both in money terms and in the amount of academics' time which it absorbs and which they could, in the absence of the RAE, spend on their research. The question then naturally arises as to whether this expensive procedure has actually improved the research produced in the UK. Strange to say this question is rarely asked. Academics

Donald Gillies

devote themselves with great energy to evaluating each other's work, but seem to be little concerned with evaluating an important government policy. Perhaps the reason for this is that it seems at first sight rather obvious that the RAE should improve the UK's research output. The procedure conforms to common sense. If we want to improve research, we should first find out who is doing good research and then give funding to the good researchers while withdrawing funding from the bad researchers. The RAE appears at first sight to be doing just this, and so the conclusion seems inevitable that introducing such a system will improve research output.

In social life, however, things are rarely simple, and judgements based on common sense can often mislead. The RAE, which is so costly in terms of money and time, is designed to improve the research output of the UK, but could it be having the opposite effect? Could it be making the research output of the UK worse instead of better? In this paper I want to argue that the RAE is indeed likely to have a negative effect on the research output of the UK. My argument will be mainly based, as the title suggests, on results which have been obtained from the study of the history and philosophy of science, and I will next consider why history and philosophy of science (henceforth HPS) is relevant to the RAE.

Let us begin with history of science. Historians of science study the great episodes of scientific advance, and the research programmes which led to exciting discoveries and to new and important knowledge. However, they also study the research programmes which failed to produce any advances, and the obstacles and difficulties which have sometimes stood in the way of scientific progress. All these matters are surely relevant to the design of a government policy intended to improve scientific research. Moreover it is not just the individual episodes which are relevant. One needs to analyse the underlying general principles which favour scientific advance, or, conversely, the general nature of the obstacles which impede scientific progress. This task of generalising from history of science falls to the practitioners of philosophy of science. Thus I think there can be little doubt that HPS is highly relevant to assessing the effectiveness of the RAE.

Of course, HPS is mainly relevant to the RAE as applied to science. For the purposes of this paper, I will take 'science' in a broad sense to include mathematics, computer science, and medicine as well as the standard natural sciences such as astronomy, physics, chemistry, etc. However, I will not include in 'science', the social sciences and the humanities. Although my focus then is on

Lessons from the History and Philosophy of Science

science in the sense just defined, and although by main arguments will be drawn from HPS, I will digress slightly in the next section by considering a branch of the humanities with which I am quite familiar—namely philosophy. Thus in the next section (section 2), I will raise the question of whether the RAE has improved philosophy in the UK. This digression will prove useful since it will suggest some methodological procedures which I will use when, in sections 3 to 7, I come to apply results from HPS to the problem of the effectiveness of the RAE.

2. Has the RAE improved philosophy in the UK?

I will begin by considering philosophy in the UK in the period 1900–1975, that is to say in the twentieth century but before the introduction of the RAE. This period of almost three quarters of a century was one of great brilliance as regards philosophy in the UK. It begins in the years before the First World War with the Cambridge school founded by Moore and Russell. Moore wrote an outstanding work on ethics while Russell made his remarkable contributions to logic and philosophy of mathematics. They were followed in the next generation by Keynes who was at this time still a philosopher rather than an economist, and who made important contributions to the philosophy of probability and induction. Wittgenstein also came to Cambridge at this time, and, working with Russell, he developed his early philosophy. After the First World War, philosophy continued to flourish at Cambridge. The prodigious Frank Ramsey died in 1930 at the age of only 26, but not before he made remarkable contributions to the philosophy of logic, probability and mathematics. In that same year, however, Wittgenstein obtained a post at Cambridge where he remained with short interruptions until 1947, rising to the position of professor. It was during this period that Wittgenstein developed his later philosophy, and wrote most of his book: *Philosophical Investigations*. Wittgenstein died in 1951, and *Philosophical Investigations* was published posthumously in 1953. Many people regard this book as the greatest philosophical masterpiece of the twentieth century.

After the Second World War, in the period 1945–75, philosophy continued to flourish in the UK, but Cambridge was no longer the main centre. Instead this period is dominated by the ordinary language school of philosophy at Oxford, and by Popper's school in London. Ordinary language philosophy had much in common with

Donald Gillies

Wittgenstein's later philosophy, but it was developed in a somewhat different way at Oxford by figures such as Ryle and Austin. Popper and his school focussed on philosophy of science. The school included not only Popper himself, but Lakatos, and, one might say, half Feyerabend. I say 'half Feyerabend' because Feyerabend started as Popper's assistant, and his first lectureship was in Bristol in the UK. However, he later worked in the USA and continental Europe. Still up to the death of Lakatos in 1974, Feyerabend visited London frequently giving regular lecture courses there.

The Oxford school and Popper's school were not on the best of terms. Their feud was foreshadowed by a famous argument between Popper and Wittgenstein which took place in 1946, and has become the subject of Edmonds and Eidinow's interesting 2001 book: *Wittgenstein's Poker*. However, a feud between two rival and very different schools of philosophy is surely yet another sign that philosophy was flourishing in the UK.

Moreover in the period 1945 to 1975, there were other significant developments in philosophy in the UK which lay somewhat outside the main schools just described. Perhaps the most important of these was the beginning of the philosophy of artificial intelligence. In his classic article published in *Mind* in 1950, Turing argued that computers could eventually equal if not surpass human beings in intellectual skills. Lucas replied in 1961 with a famous argument which used Gödel's incompleteness theorems to try to show that the human mind would always remain ahead of any possible computer.

So surveying philosophy in the UK in the years 1900–1975 one cannot but conclude that this was a brilliant period in which philosophy of the highest quality was produced in abundance in the UK. However, this was all before anyone had thought of introducing the RAE, and there was no system in operation which resembled the RAE in any way. Doesn't this begin to suggest that the RAE may be quite unnecessary?

At this point, however, a stern defender of the RAE might say: 'This is all very well, but there is never any good reason for being complacent. Even if things are going well, there is still always room for improvement, and the RAE was designed to achieve such improvement.' Well, the RAE may have been designed to bring about improvement, but did it succeed? The RAE has been with us since 1986. Has the philosophical output of the UK improved during these nearly 21 years?

Having formulated this last question, I have to confess that I am unable to answer it. I am genuinely unsure as to whether

Lessons from the History and Philosophy of Science

philosophy produced in the UK in the last 10 years say has been good or not. Naturally some pieces of UK philosophy have struck me as good, while others have appeared to me as quite bad. However, even these judgements I hold tentatively and with considerable uncertainty, while I am, at the same time, completely certain that others hold quite different opinions from mine about what is good and what is bad. Earlier I talked about philosophy in the period 1945 to 1975, and here I felt confident about the opinions I expressed. After the lapse of thirty years, a historical perspective has been obtained and it becomes easy to judge who were the important philosophers, and to evaluate the significance of their contributions. The situation is quite otherwise with contemporary philosophy where it is hard to say which philosophical works have a real importance, which are now esteemed only because of some passing fashion, and which are at present unjustly neglected because their true significance has not yet been grasped.

Just as I have been writing about philosophy in the UK in the period 1945 to 1975 which ended thirty years ago, let us imagine someone writing about the period 1986 to 2007 in the year 2037. Can we guess what such a writer might say? There could be a spectrum of judgement along a scale running from those most favourable to the RAE to those most hostile to it. A very favourable judgement might go like this: 'Although few realised it at the time, the year 1986 proved a turning point for philosophy in the UK. The introduction of the RAE in that year stimulated the philosophy community in the UK to greater efforts and we can now recognize that the works they produced in the subsequent decades outshone those completed in the first three quarters of the twentieth century.' A very hostile judgement might go like this: 'There was a brilliant flowering of philosophy in the UK in the first three quarters of the twentieth century. Unfortunately after about 1975, a decline set in and the situation was made worse by the introduction of the RAE in 1986. In the decades following this introduction, philosophy in the UK, which earlier in the century had been so brilliant, sank to a very low point.' These two judgements could be taken as marking the extreme points of a scale. Where along this scale will the majority of writers in 2037 be found? I really would not like to say.

One important point has emerged from the preceding discussion, namely that it is often very difficult to judge the quality of contemporary research. It is often only after the elapse of a considerable interval of time—say thirty years or more—that one can say with any confidence that a piece of research was either

Donald Gillies

genuinely good or really very bad. The situation is indeed worse than this, for, as we shall see later in the paper, contemporary judgements on the quality of pieces of research, are often wildly mistaken in the light of what emerges later on. But this exposes a key weakness in the RAE. The RAE naturally relies on contemporary judgements as to which researchers are good and which are bad, but such judgements are difficult to make and may often be found to be quite wrong in the light of later developments.

I will pick up this point again in what follows, but I would like now to consider another approach to assessing whether the effect of the RAE on philosophical output is likely to be good or bad. As we have seen the direct approach of asking whether the introduction of the RAE improved the quality of philosophy in the UK did not yield any very clear answer. This is why I suggest using another approach which I call the 'counterfactual methodology'. The idea here is to consider research carried out in the past, and ask whether, if the RAE had been in existence at that time, it would have improved that research or would, on the contrary, have made it worse.

Let us apply this counterfactual methodology to philosophy in the UK in the 1930s. Suppose the RAE had been introduced in say 1936, what effect would it have had on philosophy in the UK. This was the time when Wittgenstein was developing his later philosophy at Cambridge and writing early drafts of what later became his *Philosophical Investigations*, judged by many to be the greatest philosophical book of the twentieth century. How would Wittgenstein have fared in a RAE conducted according to the existing rules?

Actually we can answer this question quite easily. Wittgenstein was offered a position at Cambridge in 1930, rose to becoming professor there, and resigned from his chair in 1947. During these 17 years he published nothing. In fact the last philosophical work which he published in his lifetime was a paper entitled: 'Some Remarks on Logical Form' which appeared in the *Journal of the Aristotelian Society* in 1929. Wittgenstein had agreed to give a talk to the Aristotelian Society that year. The Aristotelian Society insists that papers are always printed in advance, and this is why Wittgenstein's paper was published. However Wittgenstein decided shortly after the printing had taken place that the paper was worthless, and at the meeting, actually talked on another topic (see Monk, 1990, 272–3 for details). After this experience, Wittgenstein

Lessons from the History and Philosophy of Science

became very reluctant to publish anything which he had not worked on for a long period, and this explains why he published nothing further for the next 17 years.

Now what happens under the RAE rules to academics who publish nothing? They are classified as research inactive, and their fate is not agreeable. Their research time is removed, and they have to spend more time teaching. Moreover they are at risk of being sacked. If the RAE had been in existence in 1936, and the rules had been applied without fear or favour, then this is the fate which would have overtaken Wittgenstein.

Now a defender of the RAE might at this point object to my analysis on the following grounds: 'Wittgenstein published nothing between 1930 and 1947 because he was under no pressure to do so. Had the RAE been introduced in 1936, he would certainly have 'knuckled under' and published some stuff.' Unfortunately for this argument, numerous memoirs and the magnificent (1990) biography of Wittgenstein by Ray Monk have given us quite a vivid picture Wittgenstein's character, and this leaves no doubt that Wittgenstein was the last person on earth to have 'knuckled under' and obeyed the directives of the RAE.

In fact Wittgenstein, despite his great intellectual brilliance, seems to have disliked the company and habits of academics and to have preferred associating with simple folk. Karl Britton, a former student of Wittgenstein's, very clearly describes this attitude of the master (quoted from Pitcher, 1964, 12):

He had, he said, only once been to high table at Trinity and the clever conversation of the dons had so horrified him that he had come out with both hands over his ears. The dons talked like that only to score: they did not even enjoy doing it. He said his own bedmaker's conversation, about the private lives of her previous gentlemen and about her own family, was far preferable: at least he could understand why she talked that way and could believe that she enjoyed it.

As a result of these attitudes, Wittgenstein showed a strong propensity to abandon seats of academic learning, and go off to remote spots in the country where he could associate with simple country folk. This propensity manifested itself as early as 1913, where he decided to go off to live alone in a remote area of Norway for two years. Russell tried in vain to dissuade him, and wrote about it in a letter as follows (quoted from Monk, 1990, 91):

Donald Gillies

I said it would be dark, & he said he hated daylight. I said it would be lonely, & he said he prostituted his mind talking to intelligent people. I said he was mad & he said God preserve him from sanity. (God certainly will.)

This episode illustrates the extreme obstinacy and determination of Wittgenstein's character. He was really not the sort of man who would have been prepared to 'knuckle under' and obey some government regulation which he regarded as mistaken.

Wittgenstein went to Norway but did not stay for two years because of the outbreak of the First World War. After the War, despite having become famous in philosophical circles because of the publication of his *Tractatus*, he decided to give up philosophy and worked as a schoolmaster in remote Austrian villages between 1920 and 1926. He refused to attend any of the meetings of the Vienna Circle which greatly admired his work. However Wittgenstein was eventually persuaded to return to academic life in Cambridge in 1930. Yet he remained full of longings for a simple life of manual toil in some remote country location. He even applied in 1935, at the height of Stalinism, to work as a labourer on a collective farm in Russia. Perhaps luckily for him the Russians turned down his application (see Monk, 1990, 351). So Wittgenstein went back to his hut in Norway for a year instead.

Many people might regard the job of being professor of philosophy at Cambridge as rather an agreeable one, but not so Wittgenstein. In a letter to Malcolm in 1945, Wittgenstein wrote: '... the absurd job of a prof. of philosophy ... is a kind of living death' (quoted from Malcolm, 1958, 38). At this time he was contemplating resigning his professorship at Cambridge, which he did in 1947. Just before his resignation, Wittgenstein wrote (quoted from Monk, 1990, 516):

Cambridge grows more and more hateful to me. The disintegrating and putrefying English civilization. A country in which politics alternates between an evil purpose and *no* purpose.

After resigning his chair, Wittgenstein went off in 1948 to live in a remote country district in Galway on the west coast of Ireland.

These episodes give a vivid illustration of Wittgenstein's character and tastes. In the light of these, is it possible that, if the RAE had been introduced in 1936, he would have agreed to its demands and started publishing some of his work? I find it quite inconceivable that he would have done so. Malcolm in his Memoir records (49) that in the academic year 1946–7, Wittgenstein stated

Lessons from the History and Philosophy of Science

that 'he was not going to be 'stampeded' into publishing prematurely.' In fact he had published nothing for over 17 years at that stage of his career.

If there is still any doubt on this point, it could be added that Wittgenstein was also highly contemptuous of the typical academic procedures which are enshrined in the RAE. This is illustrated vividly in letters written by Wittgenstein to Malcolm in 1945 and 1948. It should be explained that Wittgenstein was very fond of reading American detective magazines—particularly those published by Street and Smith. In the 1930s and 1940s, *Mind* was a leading English philosophy journal, as indeed it still is today. To have a series of papers published in *Mind* would be regarded as a strong point in favour of any researcher according to the usual RAE criteria. Wittgenstein, however, far from endorsing these RAE criteria is very sarcastic about them, and compares *Mind* unfavourably with the detective magazines of Street and Smith (cf. Malcolm, 1958, 32). In 1945 he wrote to Malcolm:

If I read your mags I often wonder how anyone can read 'Mind' with all its impotence & bankruptcy when they could read Street & Smyth mags. Well, everyone to his taste.

In another letter to Malcolm in 1948, he elaborated the comparison:

Your mags are wonderful. How people can read Mind if they could read Street & Smith beats me. If philosophy has anything to do with wisdom there's certainly not a grain of that in Mind, & quite often a grain in the detective stories.

Suppose then that the RAE had been introduced in 1936. Are we seriously to suppose that Wittgenstein would have 'knuckled under' and submitted papers for publication in *Mind*? Given his character and views, it is altogether out of the question that he would have done so. His reaction is entirely predictable. In the face of such a demand, he would have undoubtedly have left Cambridge in disgust and gone off to his hut in Norway.

The effect of the introduction of the RAE in 1936 would then have been to hound Wittgenstein out of Cambridge. Hardly a result which should increase our confidence in the merits of the RAE! What actually happened was that Wittgenstein was offered a Chair in Philosophy at Cambridge in 1939, despite having published nothing for ten years. Such an appointment would of course be almost impossible under a RAE regime. Even if the members of the appointments panel were sympathetic to a candidate who had

Donald Gillies

published nothing for ten years, they could hardly overlook the fact that such a professor would contribute nothing to the RAE, and would indeed set a bad example to the rest of the department. So if the RAE had been introduced in 1936, Wittgenstein would have been very unlikely to have become Professor of Philosophy at Cambridge.

Wittgenstein was perhaps rather excessively reluctant to publish, but how can we condemn his strategy in general terms? Wittgenstein was not of course really research inactive while at Cambridge. Although he published nothing between 1929 and 1951, he produced roughly thirty thousand pages of notebooks, manuscripts, and typescripts on philosophy in that period (Malcolm, 1958, 84). That is an average rate of about 26 pages a week. Wittgenstein's view was that he shouldn't publish anything until he had thought and rethought about it, and worked through it many times revising and correcting. He believed that only in this way could he produce philosophical work of lasting value. Now how can we say he was wrong about this? After all, his strategy worked. At the end of his long years of rethinking, revising and correcting, he produced a book (*Philosophical Investigations*) which many regard as the philosophical masterpiece of the twentieth century.

I am not saying that every philosopher should adopt Wittgenstein's strategy. Other philosophers work in a quite different way and yet produce just as good philosophy. It is partly a matter of style and temperament. Russell, for example, who was in my opinion just as good a philosopher as Wittgenstein, worked in quite a different way. He had no inhibitions about publishing, and, when thinking about a problem, would often publish in rapid succession a series of papers considering different solutions before finally settling on a particular approach. But, although Wittgenstein's way of working is not the only one, it is certainly a possible way of working which has produced great philosophy. It is thus obviously wrong for the RAE to rule out this strategy of delaying publication, and this is a great weakness of the whole system.

Let us now consider what response a defender of the RAE might give to this objection. He or she might reply along the following lines: 'Philosophy is a peculiar intellectual discipline, and tends to attract peculiar people. Even by the standards of philosophers, Wittgenstein was exceptionally strange. Now if we turn from philosophy to more serious intellectual disciplines such as mathematics, medicine, physics or astronomy, we shall find that these scientific disciplines are carried out by more serious people

Lessons from the History and Philosophy of Science

for whom the criteria of the RAE are certainly appropriate.’ To meet this challenge, we must turn to a consideration of science, and this therefore is a good point at which to begin the main line of argument of the paper.

In fact we will discover that many of the great scientific innovators had personalities which were no less unusual than Wittgenstein’s. It will also emerge that Wittgenstein had some advantages which several of those who made great advances in science lacked. Wittgenstein’s work was recognised very early on by individuals such as Russell and Keynes who could exercise a powerful influence in the academic world. Some other notable pioneers had the less agreeable experience of finding that their innovative work was not recognised by anyone, and indeed was rejected as absurd by those in powerful academic positions.

In considering how the RAE might affect research in science in the UK, I will apply the ‘counterfactual methodology’ introduced by the case of Wittgenstein. I will consider a number of great advances in science which occurred in the past, and ask whether, if the RAE had existed in those days, it would have helped or hindered that advance. The result of the cases which I will consider is the same as the result in the case of Wittgenstein—namely that the RAE, if it had been in existence, would constitute an obstacle to the advance.

I have chosen three cases which are designed to cover a range of different sciences. The first is in mathematics, the second in medicine, and the third in astronomy. I have chosen cases where a very striking scientific advance was made at a theoretical level. I do not, however, want to focus on theory and neglect practical applications. It is now generally agreed that the development of new technological applications of science is very important in order to make the UK competitive in the era of globalisation. I have therefore chosen three theoretical advances which had very important wealth-generating applications.

3. First Case-History: Frege and Mathematical Logic

My first example is taken from the field of mathematics and I want to consider an important advance made in a branch of the subject known as mathematical logic. This advance was made by Frege in a booklet published in 1879, and which is usually referred to by its German title of *Begriffsschrift*, which means literally: ‘concept-writing’. It might be objected to this example that Frege was a

Donald Gillies

philosopher rather than a mathematician. It is true that Frege wrote some very important works on philosophy, but that does not make him any less a mathematician. Other famous mathematicians such as Descartes and Leibniz also wrote on philosophy. Frege worked all his life in the mathematics department of Jena university. The *Begriffsschrift* does contain some interesting philosophical remarks, but it is mainly formal in character. Its contribution is to what is now called mathematical logic and it is difficult to deny that mathematical logic is a branch of mathematics.

Indeed Frege's *Begriffsschrift* may justly be said to have introduced modern mathematical logic. In this work Frege presents for the first time an axiomatic-deductive development of the propositional calculus and of the predicate calculus (or quantification theory). The propositional and predicate calculi are the first things introduced in any modern treatment of mathematical logic. What is still more surprising is that the expositions of these calculi in contemporary textbooks are often quite close to the original expositions of Frege. Two well-known and widely used textbooks of mathematical logic are Mendelson (1964) and Bell and Machover (1977). Mendelson introduces the propositional calculus and quantification theory in chapters 1 & 2, while Bell and Machover introduce them in chapters 1, 2 & 3. Of course they both give many results and approaches which were discovered after Frege, but they do also give an axiomatic-deductive treatment which has a lot in common with Frege's and indeed uses some of the same axioms that Frege used.¹ Frege's treatment in the *Begriffsschrift* includes what is known as higher-order logic, whereas modern treatments usually limit themselves to first-order logic. However, leaving this subtlety aside we can say that Frege's treatment of both the propositional and predicate calculi is complete from a modern point of view, though his axiomatic presentation was subsequently simplified by reducing the number of axioms. Thus Frege created in the *Begriffsschrift* a whole new formal theory which is still today taken as the core of mathematical logic.

Frege's remarkable achievement has been fully recognised by experts in the field since the 1950s. In Appendix II to his English translation of the *Begriffsschrift*, Bynum very usefully collects together some evaluations by well-known scholars writing in the

¹ A detailed comparison of the *Begriffsschrift* with the treatment of the corresponding material in Mendelson (1964) and Bell and Machover (1977) is to be found in Gillies, 1992, 275–6.

Lessons from the History and Philosophy of Science

1950s and 1960s. Here are some extracts from the passages he gives. They are all quoted from Bynum, 1972, 236–8.

Quine, 1952 (236): ‘... the logical renaissance might be identified with the publication of Frege’s *Begriffsschrift* in 1879 ... 1879 did indeed usher in a renaissance, bringing quantification theory and therewith the most powerful and most characteristic instrument of modern logic ... with the aid of quantification theory modern logicians have been able to illuminate the mechanism of deduction in general, and the foundations of mathematics in particular, to a degree hitherto undreamed of.’

Dummett, 1959 (238): ‘There can be no doubt that Boole deserves great credit for what he achieved ... however ... Boole cannot correctly be called “the father of modern logic”. *The* discoveries which separate modern logic from its precursors are of course the use of quantifiers ... and a concept of a formal system, both due to Frege and neither present even in embryo in the work of Boole.’

Bochenski, 1962 (237): ‘Among all these logicians, Gottlob Frege holds a unique place. His *Begriffsschrift* can only be compared with one other work in the whole history of logic, the *Prior Analytics* of Aristotle. The two cannot quite be put on a level, for Aristotle was the very founder of logic, while Frege could as a result only develop it. But there is a great likeness between these two gifted works.’

William and Martha Kneale, 1962 (236–7): ‘Frege’s *Begriffsschrift* is the first really comprehensive system of formal logic... . Frege’s work ... contains all the essentials of modern logic, and it is not unfair either to his predecessors or to his successors to say that 1879 is the most important date in the history of the subject.’

Frege carried out his researches in mathematical logic for purely theoretical reasons, but, as so often happens, his results turned out to be of great practical importance. Mathematical logic is one of the fundamental tools of present-day computer science, and one can further say that the computer as we know it today could not have developed without a prior development of mathematical logic. Detailed accounts of the use of mathematical logic in computer science and in the development of computing are contained in Davis (1988a & b) and in Gillies (2002). There are many specific examples of the application of mathematical logic in computer

Donald Gillies

science, but at a very fundamental level one can say that the *Begriffsschrift* is the first example of a fully formalised language, and so, in a sense, the precursor of all programming languages (see Davis, 1988b, 316).

Thus Frege's research turned out to provide some of the fundamental tools for a wealth-generating technological advance. Consequently Frege's research work must be the kind of research work which a nation like the UK should try to encourage. This brings us to the question of whether Frege's research would have been helped if, counterfactually, there had been a RAE regime operating in Germany in his day. Suppose there had been a German RAE in the 1880s, how would Frege have done? The answer is: 'not very well.'

In Appendix I to his translation of the *Begriffsschrift*, Bynum gives in full the contemporary reviews of the work, all written in the years 1879 and 1880. It is very interesting to compare these with the evaluations of the same work made with the benefit of historical perspective in the 1950s and the 1960s. These are given by Bynum in his Appendix II, and we have already quoted some passages.

Turning now to the contemporary reviews of the *Begriffsschrift*, they were 6 in number—all quotations from them will be from the versions in Bynum, 1972, 209–35. Four were written by Germans, one by a Frenchman (Tannery) and one by an Englishman (Venn). Only one of these reviews, which was written by a German: Lasswitz, is favourable. The other three German reviews do make some favourable remarks, but one cannot help wondering whether these are designed to be polite to a compatriot and colleague, since they are contradicted by the majority of the detailed comments on the work which are highly unfavourable. Thus Hoppe concludes his review by saying (210): 'On the whole, the book, as suggestive and pioneering, is worth while.' However earlier in the same review he had written (209): '... we doubt that anything has been gained by the invented formula language itself.' Similarly Michaelis concludes his review (218): 'His work ... certainly does not lack importance.' However, this rather contradicts the following harsh judgement given in the body of the review, where Michaelis says (217): '... Frege has to pass over many things in formal logic and detract even more from its content... . The content of logic which has been much too meagre up to now, should not be decreased, but increased.' In contrast to the later critics who saw Frege as having made an enormous step forward in logic, Michaelis actually thinks that Frege has decreased, or detracted from, the content of logic.

Lessons from the History and Philosophy of Science

However, the harshest German review comes from the most famous German logician of the time: Schröder. Schröder actually upbraids Lasswitz for having written a review supporting the *Begriffsschrift*, and says of this review (220) that he casts ‘a disapproving glance at it’. He refers to Lasswitz later (221) as ‘the Jena reviewer’, which seems to imply that Lasswitz’s favourable judgement arises from some personal connection with Frege. Schröder’s own judgement on the *Begriffsschrift* is very negative indeed. He thinks that Frege has done nothing which has not already been done much better by other people. As he says (220): ‘... the present little book makes an advance which I should consider very creditable, if a large part of what it attempts had not already been accomplished by someone else, and indeed (as I shall prove) in a doubtlessly more adequate fashion.’ It soon becomes clear that this other person is Boole. Indeed Schröder goes on to say (221) that, leaving aside the question of function and generality and some applications, ‘... the book is devoted to the establishment of a formula language, which essentially coincides with Boole’s mode of presenting *judgements* and Boole’s calculus of judgments, and which certainly in no way achieves more.’ Here Schröder does seem to make an exception in favour of Frege’s treatment of generality but this appearance is deceptive for he later goes on to say that Frege’s treatment of generality is in no way superior to the Boolean. He writes (229–30): ‘Now in the section concerning “generality”, Frege correctly lays down stipulations that permit him to express such judgements precisely. I shall not follow him slavishly here; but on the contrary, show that one may not perchance find a justification here for his other deviations from Boole’s notation, and the analogous modification or extension can easily be achieved in Boolean notation as well.’ (Logicians will at once see from this that Schröder has completely failed to grasp the importance of introducing the quantifiers.) But could Frege at least be defended on the grounds that he has shed some light on the logical nature of *arithmetical* judgements? ‘Not so’, argues Schröder, ‘for that matter too has already been cleared up by someone else.’ In his own words (231): ‘According to the author, he undertook the entire work with the intention of obtaining complete clarity with regard to the logical nature of *arithmetical* judgements, and above all to test “how far one could get in arithmetic by means of logical deductions alone”. If I have properly understood what the author wishes to do, then this point would also be, in large measure, already settled—namely, through the perceptive investigations of Hermann Grassmann.’ After dismissing Frege’s work so completely, it is rather surprising

Donald Gillies

that Schröder concludes (231): ‘May my comments, however, have the over-all effect of encouraging the author to further his research, rather than discouraging him.’ Perhaps Schröder felt some pangs of guilt about writing so harshly about the work of a young researcher in his field. The two non-German reviews of the *Begriffsschrift* are if anything even more dismissive than the German reviews, and contain no favourable remarks at all. Tannery in France writes (233): ‘In such circumstances, we should have a right to demand complete clarity or a great simplification of formulas or important results. But much to the contrary, the explanations are insufficient, the notations are excessively complex; and as far as applications are concerned, they remain only promises.’ Nowadays one of Frege’s great advances is considered to be the replacement of the Aristotelian analysis in terms of subject and predicate by an analysis using function and argument. Tannery notes this change but regards it as a mistake (233): ‘The [author] abolishes the concepts of *subject* and *predicate* and replaces them by others which he calls *function* and *argument*... . We cannot deny that this conception does not seem to be very fruitful.’ Finally Venn in England entirely agrees with Schröder that Frege has made no advance over Boole and has indeed taken a step backwards. Venn writes (234): ‘... it does not seem to me that Dr. Frege’s scheme can for a moment compare with that of Boole. I should suppose, from his making no reference whatever to the latter, that he has not seen it, nor any of the modifications of it with which we are familiar here. Certainly the merits which he claims as novel for his own method are common to every symbolic method.’ Venn, moreover, has no kind words at the end of his review, but concludes by saying (235): ‘... Dr Frege’s system ... seems to me cumbrous and inconvenient.’ It is worth noting here that Frege’s advances over Boole which seem so obvious today and which are mentioned by Dummett in the passage quoted above, were not appreciated at all by Schröder and Venn—two of the leading logicians of Frege’s time. So to sum up. If we go carefully through the six contemporary reviews of Frege’s *Begriffsschrift*, we find only one which takes a positive view of Frege’s work. In the other five, there is a consensus to be found that the *Begriffsschrift* makes no advance on what has already been done, particularly by Boole and the Booleans, and indeed that it is in many respects inferior to and a step back from already existing logical works.

What is remarkable is that Frege was not discouraged by these damning reviews, but continued his work on his logicist programme for the next 24 years. However, his subsequent books were, if

Lessons from the History and Philosophy of Science

anything, even less successful than the *Begriffsschrift*. The *Foundations of Arithmetic* published in 1884 received only 3 reviews—all unfavourable. In 1891 Frege wanted to publish a third book in the series, but, perhaps not surprisingly, found it hard to find a publisher. Eventually, however, (Bynum, 1972, 34): ‘the publisher Hermann Pohle in Jena ... agreed to print the book in two instalments, the publication of the second part to be dependent upon a good reception of the first. So, in late 1893, the first volume of *The Basic Laws of Arithmetic* appeared.’ This book got only two reviews—both unfavourable. In the light of this Frege had to publish the second volume which appeared in 1903, at his own expense. In the 1890s and 1900s a few avant-garde researchers—notably Peano and Russell—did begin to study and develop Frege’s ideas. However, even when Frege retired from Jena at the age of 70 in 1918, general recognition had still eluded him. The situation was well summed-up by Bochenski in 1962 (quoted from Bynum, 1972, 237–8):

It is a remarkable fact that this logician of them all had to wait twenty years before he was at all noticed, and another twenty before his full strictness of procedure was resumed by Lukasiewicz. In this last respect, everything published between 1879 and 1921 fell below the standard of Frege, and it is seldom attained even today.

Anyone who is concerned with formulating policies concerned with research, should in my opinion read carefully the two appendices to Bynum, 1972. They amount to only 30 pages, but they demonstrate in a conclusive fashion that the method of peer review can, in some cases, go very wrong. It does happen that the majority of contemporary researchers in a field can judge as worthless a piece of research which is later, with the benefit of historical perspective, seen as constituting a major advance.

Now the RAE does clearly rely on peer review because the value of each researcher is judged by a committee of experts in the field. Indeed the RAE in a sense involves a double use of peer review, because the members of the RAE consider only research which has been published, and, to get a piece of work published, a researcher has usually to submit it to a journal which uses peer review to assess whether it is worth publishing. The problem facing those like Frege, whose work is judged of little value by the majority of their peers, is that they may find it difficult to publish at all. This applied to Frege himself. He wrote two papers replying to the criticism of Schröder and Venn that his work was inferior to that of Boole.

Donald Gillies

However he was unable to publish these papers (cf. Bynum, 1972, 21), and they only appeared long after Frege's death. Those who find their peers against their work will certainly be excluded from publishing in the more famous journals and may have to resort to publishing in lower quality journals or even to publishing the material in book form at their own expense, as Frege did in 1903. Now theoretically the RAE committee reads carefully and judges on their merits all the works submitted to it, but of course in practice papers which have appeared in high ranking journals, or books which have been published by prestigious firms such as Oxford University Press, are likely to be judged more favourably. Conversely papers which have appeared in low ranking journals, or, worse still, books published at the author's own expense—something usually called 'vanity publishing'—are likely to be judged more harshly. We have therefore to conclude that if the RAE had existed in Germany in the 1880s Frege would have got a very low rating.

Even in the RAE free Germany of the period, Frege did not have an easy time. A fascinating portrait of him in the years 1910–14 is given by Carnap in his intellectual autobiography (Carnap, 1963). Carnap's involvement with Frege appears to have come about rather by chance. Carnap's family lived in Jena and Carnap went to the local university where Frege taught. Carnap writes (1963, 5):

In the fall of 1910, I attended Frege's course "Begriffsschrift" (conceptual notation, ideography), out of curiosity, not knowing anything either of the man or the subject except for a friend's remark that somebody had found it interesting. We found a very small number of other students, there. Frege looked old beyond his years. He was of small stature, rather shy, extremely introverted. He seldom looked at the audience. Ordinarily we saw only his back, while he drew the strange diagrams of his symbolism on the blackboard and explained them. Never did a student ask a question or make a remark, whether during the lecture or afterwards. The possibility of a discussion seemed to be out of the question.

Earlier in his account, Carnap says (1963, 4):

Gottlob Frege (1848–1925) was at that time, although past 60, only Professor Extraordinarius (Associate Professor) of mathematics in Jena. His work was practically unknown in Germany; neither mathematicians nor philosophers paid any attention to it.

Lessons from the History and Philosophy of Science

It was obvious that Frege was deeply disappointed and sometimes bitter about this dead silence.

Carnap, however, took a liking to Frege's work and attended his two advanced courses "Begriffsschrift II" in 1913 and his course *Logik in der Mathematik* in 1914. Carnap records that "Begriffsschrift II" was attended by 3 students: Carnap, a friend of Carnap's, and (Carnap, 1963, 5) 'a retired major of the army who studied some of the new ideas in mathematics as a hobby.'

We can see from this that Frege's career was hardly a great success, but, if there had been a RAE regime in Germany, things would have gone even worse for him. As we have seen, Frege would undoubtedly have got a low rating in the RAE exercise, and the inevitable penalties would have fallen on his head. His research time would have been cut and he would have been forced to take on extra teaching duties. Thus he would not have had the necessary research time to develop his mathematical logic. Moreover, as we can see from Carnap's description, Frege may not have performed particularly well as a teacher. He seemed to attract very few students, and his teaching technique does not appear to have been of the kind recommended by educational experts. Having failed as both a researcher and a teacher, there is little doubt than, under a RAE regime, Frege would have been forced to retire early rather than allowed to stay on until he was 70. Thus Carnap would never have been able to attend his lectures, and the development and diffusion of the new important ideas of mathematical logic would have been held up still further.

4. Second Case-History: Semmelweis and Antisepsis

My second case-history, as we shall see, has many points in common with the first. However, it does differ very strikingly as regards the branch of science in which the research was conducted. Frege's research was purely theoretical, and was carried out in a branch of mathematics, mathematical logic, which is closely linked to philosophy. Semmelweis's research by contrast was highly empirical, and was carried out in medicine. In fact Semmelweis's investigation was into the causes of a terrible disease (puerperal fever) which affected women who had just given birth. Puerperal fever was, at the time, the principal cause of death in childbirth.

Semmelweis was Hungarian, but studied medicine at the University of Vienna. In 1844 he qualified as a doctor, and, later in

Donald Gillies

the same year obtained the degree of Master of Midwifery. From then until 1849, he held the posts of either aspirant to assistant or full assistant at the first maternity clinic in Vienna. It was during this period that he carried out his research.²

The Vienna Maternity Hospital was divided into two clinics from 1833. Patients were admitted to the two clinics on alternate days thereby producing, unintentionally, a system of random allocation. Between 1833 and 1840, medical students, doctors and midwives attended both clinics, but, thereafter, although doctors went to both clinics, the first clinic only was used for the instruction of medical students who were all male in those days, and the second clinic was reserved for the instruction of midwives. When Semmelweis began working as a full assistant in 1846, the mortality statistics showed a strange phenomenon

Between 1833 and 1840, the death rates in the two clinics had been comparable, but, in the period 1841–46, the death rate in the first clinic was 9.92% and in the second clinic 3.88%. The first figure is more than 2.5 times the second—a difference which is certainly statistically significant. The quoted figures actually underestimate the difference since some severe cases of puerperal fever were removed from the first clinic to the general hospital where they died—thereby disappearing from the first clinic's mortality statistics. This rarely happened in the second clinic. Semmelweis was puzzled and set himself the task of finding the cause of the higher death rate in the first clinic.

Semmelweis followed a procedure rather similar to Popper's conjectures and refutations. He considered in turn a number of hypotheses as to what might be the cause of the difference between the two clinics. He then compared these hypotheses to the facts, and found that each one of a long series of hypotheses was refuted by this comparison. Eventually, however, Semmelweis did hit on a hypothesis which was corroborated by the observations.

The first hypothesis considered by Semmelweis was that the higher death rate in the first clinic was due to 'atmospheric-cosmic-terrestrial' factors. This sounds strange but is just a way of referring to the miasma theory of disease which was standard at the time. However Semmelweis pointed out that it could not explain the

² This account of Semmelweis's research is a shortened version of the one given in my paper: Gillies (2005). That paper also contains more detailed references to the considerable literature on Semmelweis. Semmelweis's own account of his researches in Semmelweis (1861) is also worth consulting.

Lessons from the History and Philosophy of Science

different mortality rates in the first and second clinics. These were under the same roof and had an ante-room in common. So they must be exposed to the same 'atmospheric-cosmic-terrestrial' influences. Yet the death rates in the two clinics were very different.

The next hypothesis was that overcrowding was the key factor, but this too was easily refuted since the second clinic was always more crowded than the first, which, not surprisingly had acquired an evil reputation among the patients, almost all of whom tried to avoid it.

In this sort of way Semmelweis eliminated quite a number of curious hypotheses. One concerned the appearance of a priest to give the last sacrament to a dying woman. The arrangement of the rooms meant that the priest, arrayed in his robes, and with an attendant before him ringing a bell had to pass through five wards of the first clinic before reaching the sickroom where the woman lay dying. The priest had, however, direct access to the sickroom in the case of the second clinic. The hypothesis then was that the terrifying psychological effect of the priest's appearance debilitated patients in the first clinic, and made them more liable to puerperal fever. Semmelweis persuaded the priest to come by a less direct route, without bells, and without passing through the other clinic rooms. The two clinics were made identical in this respect as well, but the mortality rate was unaffected.

After trying out these hypotheses and others unsuccessfully, Semmelweis was in a depressed state in the winter of 1846–7. However a tragic event early in 1847 led him to formulate a new hypothesis. On 20th March 1847, Semmelweis heard with sorrow of the death of Professor Kolletschka. In the course of a post-mortem examination, Professor Kolletschka had received a wound on his finger from the knife of one of the students helping to carry out the autopsy. As a result Kolletschka died not long afterwards of a disease very similar to puerperal fever. Semmelweis reasoned that Kolletschka's death had been owing to cadaverous matter entering his bloodstream. Could the same cause explain the higher death rate of patients in the first clinic? In fact professors, assistants and students often went directly from dissecting corpses to examining patients in the first clinic. It is true that they washed their hands with soap and water, but perhaps some cadaverous particles still adhered to their hands. Indeed this seemed probable since their hands often retained a cadaverous odour after washing. The doctors and medical students might then infect some of the patients in the first clinic with these cadaverous particles, thereby giving them

Donald Gillies

puerperal fever. This would explain why the death rate was lower in the second clinic, since the student midwives did not carry out post-mortems.

In order to test this hypothesis, Semmelweis, from some time in May 1847, required everyone to wash their hands in disinfectant before making examinations. At first he used *chlorina liquida*, but, as this was rather expensive, chlorinated lime was substituted. The result was dramatic. In 1848 the mortality rate in the first clinic fell to 1.27%, while that in the second clinic was 1.30%. This was the first time the mortality rate in the first clinic had been lower than that of the second clinic since the medical students had been divided from the student midwives in 1841.

Through a consideration of some further cases, Semmelweis extended his theory to the view that, not just cadaverous particles, but any decaying organic matter, could cause puerperal fever if it entered the bloodstream of a patient.

Let us now look at Semmelweis's theory from a modern point of view. Puerperal fever is now known as 'post-partum sepsis' and is considered to be a bacterial infection. The bacterium principally responsible is *streptococcus pyogenes*, but other *streptococci* and *staphylococci* may be involved. Thus, from a modern point of view, cadaverous particles and other decaying organic matter would not necessarily cause puerperal fever but only if they contain a large enough quantity of living *streptococci* and *staphylococci*. However as putrid matter derived from living organisms is a good source of such bacteria, Semmelweis was not far wrong.

As for the hand washing recommended by Semmelweis, that is of course absolutely standard in hospitals. Medical staff have to wash their hands in antiseptic soap (hibiscrub), and there is also a gelatinous substance (alcogel) which is squirted on to the hand. Naturally a doctor's hands must be sterilised in this way before examining any patient—exactly as Semmelweis recommended.

Not only are Semmelweis's views regarded as largely correct from a modern point of view, but the investigation which led him to them is now held up as model of good scientific method. In fact Hempel in his 1966 book: *Philosophy of Natural Science* gives a number of examples of what he regards as excellent scientific investigations, and the very first of these is Semmelweis's research into puerperal fever.

This then is the modern point of view, but how did Semmelweis's contemporaries react to his new theory of the cause of puerperal fever and the practical recommendations based on it?

Lessons from the History and Philosophy of Science

The short answer is that Semmelweis's reception by his contemporaries was almost exactly the same as Frege's. Semmelweis did manage to persuade one or two doctors of the truth of his findings, but the vast majority of the medical profession rejected his theory and ignored the practical recommendations based upon it. I discuss some of the detailed responses to Semmelweis in my longer paper on the subject (see Gillies, 2005, 178–9). Here I will only mention one typical reaction. After Semmelweis had made his discovery in 1848, he and some of his friends in Vienna wrote about them to the directors of several maternity hospitals. Simpson of Edinburgh replied somewhat rudely to this letter saying that its authors obviously had not studied the obstetrical literature in English. Simpson was of course a very important figure in the medical world of the time. He had introduced the use of chloroform for operations, and had recommended its use as a pain-killer in childbirth. His response to Semmelweis and his friends is very similar in character to Venn's review of Frege's *Begriffsschrift*.

In Vienna the Professor and Head of the Maternity Clinics, Johann Klein, was opposed to Semmelweis's ideas, and his opposition, and that of others, caused Semmelweis to leave Vienna in 1850. He did however get a position in a Maternity Hospital at Budapest in his native Hungary. Here he wrote up his new theory of the causes of puerperal fever, and answered the objections which had been made to it. These writings were published in book form in 1861, but once again had no success in persuading the medical profession to adopt his ideas.

Semmelweis's case is very similar to Frege's. Semmelweis, like Frege, had great difficulties, and, if counterfactually, there had been a RAE regime at the time, these difficulties would have become worse. Semmelweis's work would obviously have been judged by peer review to have no value, and his allowance of research time would have been reduced, so that he might not have had the time to write up his results in book form and to answer his critics.

The failure of the research community to recognise Semmelweis's work had of course much more serious consequences than the corresponding failure to appreciate Frege's innovations. In the twenty years after 1847 when Semmelweis made his basic discoveries, hospitals throughout the world were plagued with what were known as 'hospital diseases', that is to say, diseases which a patient entering a hospital was very likely to contract. These included not just puerperal fever, but a whole range of other unpleasant illnesses. There were wound sepsis, hospital gangrene,

Donald Gillies

tetanus, and spreading gangrene, erysipelas (or 'St. Anthony's fire'), pyaemia and septicaemia which are two different forms of blood poisoning, and so on. Many of these diseases were fatal. From the modern point of view, they are all bacterial diseases which can be conquered by applying the kind of antiseptic precautions recommended by Semmelweis.

In 1871, over twenty years after his rather abrupt reply to Semmelweis and his friends, Simpson of Edinburgh wrote a series of articles on 'Hospitalism'. These contained his famous claim, well-supported by statistics, that 'the man laid on the operating-table in one of our surgical hospitals is exposed to more chances of death than the English soldier on the field of Waterloo'. Simpson thought that hospitals infected with pyaemia might have to be demolished completely. So serious was the crisis, that he even recommended replacing hospitals by villages of small iron huts to accommodate one or two patients, which were to be pulled down and re-erected periodically. Luckily the theory and practice of antiseptics were introduced in Britain by Lister in 1865, and were supported by the germ theory of disease developed by Pasteur in France and Koch in Germany. The new antiseptic methods had become general by the mid 1880s, so that the hospital crisis was averted. All the same, the failure to recognise Semmelweis's work must have cost the lives of many patients.

In my longer paper on the Semmelweis case (Gillies, 2005, 180–1), I argue that, in the history and philosophy of science, it is customary to cite historical examples of excellent science in order to exemplify what are claimed to be good methodological principles for science. However instances in which the scientific community makes a mistake, as happened in the Semmelweis case or that of Frege, can also be valuable in suggesting new rules of practice designed to make such mistakes less likely in the future. From this point of view, the RAE is clearly a step backwards. Instead of learning from the mistakes which were made regarding Frege and Semmelweis, and introducing a system designed to make such mistakes less likely in the future, it does the opposite. If the RAE had been in existence in the days of Frege and Semmelweis, it would, as we have seen, have made their position even worse than it already was. Naturally the same will apply to any future brilliant innovators like Frege and Semmelweis who have the misfortune to be working in a RAE regime.

This point can also be made by introducing a distinction taken from the theory of statistical tests. Statistical tests are said to be liable to two types of error (Type I error, and Type II error). A

Lessons from the History and Philosophy of Science

Type I error occurs if the test leads to the rejection of a hypothesis which is in fact true. A Type II error occurs if the test leads to the confirmation of a hypothesis which is in fact false. Analogously we could say that a research assessment procedure commits a Type I error if it leads to funding being withdrawn from a researcher or research programme which would have obtained excellent results had it been continued. A research assessment procedure commits a Type II error if it leads to funding being continued for a researcher or research programme which obtains no good results however long it goes on. This distinction leads to the following general criticism of the RAE. The RAE concentrates exclusively on eliminating Type II errors. The idea behind the RAE is to make research more cost effective by withdrawing funds from bad researchers and giving them to good researchers. No thought is devoted to the possibility of making a Type I error, the error that is of withdrawing funding from researchers who would have made important advances if their research had been supported. Yet the history of science shows that Type I errors are much more serious than Type II errors. The case of Semmelweis is a very striking example. The fact that his line of research was not recognised and supported by the medical community meant that, for twenty years after his investigation, thousands of patients lost their lives and there was a general crisis in the whole hospital system.

In comparison with Type I errors, Type II errors are much less serious. The worst that can happen is that some government money is spent with nothing to show for it. Moreover Type II errors are inevitable from the very nature of research. Suppose research is required on some problem, and there are four different approaches to its solution which lead to four different research programmes. It may be almost impossible to say at the beginning which of the four programmes is going to lead to success. Suppose it turns out to be research programme number 3. The researchers on programmes 1, 2 & 4 may be just as competent and hard-working as those on programme 3, but, because their efforts are being made in the wrong direction, they will lead nowhere. Suppose programme 3 is cancelled in order to save money (Type I error), then all the money spent on research in the problem will lead nowhere. It will be a total loss. On the other hand if another programme (5) is also funded, the costs will be a bit higher but a successful result will be obtained. This shows why Type I errors are much more serious than Type II errors, and why funding bodies should make sure that some funding at least is given to every research school and approach

Donald Gillies

rather than concentrating on the hopeless task of trying to foresee which approach will in the long run prove successful.

The same analysis also shows why peer review as a system can often give wrong results. Let us return to our example of the problem being tackled by four different research programmes, of which programme number 3 ultimately proves successful. Let us suppose further (which indeed is often the case) that initially programme 3 attracts many fewer researchers than programmes 1, 2 & 4. Now it is characteristic of most researchers that they think their own approach to the problem is the correct one, and that other approaches are misguided. If a peer review is conducted by a committee whose researchers are a random sample of those working on the problem, then the majority will be working on programmes 1, 2 & 4, and are therefore very likely to give a negative judgement of the merits of programme 3. As the result of the recommendation of such a peer review, funding might be withdrawn from programme 3, and the solution of the problem might remain undiscovered for a long time.

5. Third Case-History: Copernicus and Astronomy

I now turn to my third example which I will deal with more concisely both because it is more familiar and because my general line of argument should by now have become fairly clear. However, it is worth looking at this example because it deals with yet another branch of science (astronomy) and also a different historical period.³

Copernicus (1473–1543) was born in which is now Poland and studied at Universities in both Poland and Italy. Through the influence of his uncle, he obtained the post of Canon of Frauenberg Cathedral in 1497, and held this position until his death. Copernicus' duties as canon seem to have left him plenty of time for other activities, and he seems to have devoted much of this time to developing in detail his new theory of the universe. This was published as *De Revolutionibus Orbium Caelestium*, when Copernicus was on his death bed. In the preface Copernicus states that he had meditated on this work for more than 36 years.

There is little doubt that during Copernicus' lifetime and for more than 50 or 60 years after his death, his view that the Earth

³ A more detailed account of Copernicus' work on astronomy is to be found in Kuhn (1957).

Lessons from the History and Philosophy of Science

moved was regarded as absurd, not only by the vast majority of the general public, but also by the vast majority of those who were expert in astronomy. It is significant that *De Revolutionibus* was not put on the index by the Roman Catholic Church until 1616. It was not until then that Copernicanism had sufficient adherents to be considered a threat.

Although the majority of expert astronomers of the period would have dismissed the Copernican view as absurd, a few such astronomers, notably Kepler and Galileo, did side with Copernicus and carried out researches developing his theory until, in due course, it won general acceptance by astronomers not influenced by the Roman Catholic Church's opposition.

Copernicus' research, like that of Frege and Semmelweis, had very important practical applications. Despite the Roman Catholic Church's opposition to his theory, his calculations were used in the reform of the calendar carried out by Pope Gregory XIII in 1582. Ironically the Protestant countries, whose astronomers were the first to accept the Copernican theory, rejected the Gregorian calendar for a long time on the ground that it had been introduced by the Roman Catholic Church and must presumably therefore be bad. Copernicus' theory was also used to produce improved astronomical tables. Reinhold used *De Revolutionibus* in the production of his Prutenic Tables which appeared in 1551. These were the first complete tables prepared in Europe for three centuries. In 1627, they were superseded by the Rudolphine Tables which Kepler produced using his much improved version of Copernicus's theory. The Rudolphine Tables were clearly superior to all astronomical tables in use before. Of course astronomical tables were applied in navigation, and so were an important tool for promoting the growth of European maritime trade.

Let us now once again apply our counterfactual methodology and consider how Copernicus would have been affected if, instead of being a Canon of Frauenberg, he had lived under a RAE regime. Of course it is indeed rather anachronistic to suppose that something like the RAE might have been in existence in such a distant historical period. Yet I think we can still say with some confidence that if it had existed then, it would have impacted negatively on Copernicus. Under a RAE regime, Copernicus would not have been allowed to continue his research peacefully as a Canon of Frauenberg. On the contrary, he would have been brought to account to make sure he was not wasting the tax-payers' money. In order to be allowed time to continue his research, he would have had to submit samples of his research work to a committee of

Donald Gillies

experts in the field. Now nearly all these experts, as we have already pointed out, would have judged that Copernicus' research was absurd and not worth funding. Thus Copernicus would have been sent off to a teaching university with little time for research, and would have had to devote most of his time to teaching astronomy to undergraduates. Naturally, as the syllabus would have been determined by the majority of his colleagues, he would have had to teach, not his new theory, but the standard Aristotelian-Ptolemaic account of astronomy. So Copernicus, under a system of funding of RAE type, would have been deprived of his research time, and forced to spend his days teaching the Aristotelian-Ptolemaic account of astronomy. Meanwhile the leading experts of the Aristotelian-Ptolemaic theory would have had posts at the well-funded research universities giving them plenty of time to pursue their research. No doubt they would have developed Aristotelian-Ptolemaic theory by means of ever more mathematically ingenious combinations of epicycles. It need hardly be said that all this would have acted as an extreme dampener on the progress of astronomy.

I have given three examples of cases in which a regime of RAE type would have impeded rather than helped scientific advance. Of course many more cases along the same lines could be described, but it will now be more fruitful to turn from history of science to philosophy of science. In the next section, I will try to analyse the factors, which in the cases of Copernicus, Frege and Semmelweis, led to the failure of the peer review method. As we shall see, this is not a problem which has arisen in just a few cases, but is an underlying pattern in the development of science.

6. Kuhn's Distinction between Normal and Revolutionary Science

The part of philosophy of science which I would like to consider is Kuhn's theory of scientific development as set out in his *The Structure of Scientific Revolutions* (1962). Kuhn's view is that science develops through periods of *normal science* which are characterised by the dominance of a *paradigm*, but which are interrupted by occasional revolutions during which the old paradigm is replaced by a new one. Kuhn gives three main examples of scientific revolutions. These are the Copernican

Lessons from the History and Philosophy of Science

Revolution, the Chemical Revolution, and the Einsteinian Revolution. As we have already discussed Copernicus, I will illustrate Kuhn's views by a brief account of the other two examples.

The Chemical Revolution. The main theme of the chemical revolution was the replacement of the *phlogiston* theory by the *oxygen* theory, though there were many other important changes as well. According to the phlogiston theory, bodies are inflammable if they contain a substance called phlogiston, and this is released when the body burns. The phlogiston theory was also used to explain the calcination of metals. When a metal is heated in air, in many cases it turns into a powder known as the *calx*, e.g. iron → rust. Conversely the calx is usually found in ores of the metal, and the metal itself could often be obtained by heating with charcoal. These transformations were explained by postulating that

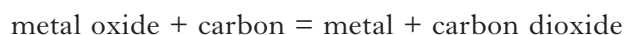


When we heat a metal, phlogiston is given off, and the calx remains. Conversely when we heat the calx with charcoal, since charcoal is very rich in phlogiston because it burns easily, the phlogiston from the charcoal combines with the calx to give the metal.

In the oxygen theory, burning is explained as the combination of the substance with oxygen; while the calx is identified with the oxide of the metal. So turning a metal into its calx by heating in air is explained by the equation



Similarly obtaining the metal by heating the calx with charcoal is explained by the equation



The oxygen theory was developed by Lavoisier. At the beginning of his researches in 1772, he was already sceptical of the then dominant phlogiston theory. In the next decade or so, many experimental discoveries concerning gases were made. These discoveries were mainly owing to the English experimental chemists—particularly Priestley and Cavendish. However, these English chemists remained faithful to the phlogiston theory. For example Priestley referred to what we now call oxygen as dephlogisticated air. Lavoisier, on the other hand, reinterpreted their results in terms of his new and developing oxygen theory. Lavoisier's new paradigm for chemistry was set out in his *Traité élémentaire de chimie* of 1789, and within a few years it was adopted

Donald Gillies

by the majority of chemists. Priestley, however, who lived until 1804 never gave up the phlogiston theory.

The Einsteinian Revolution. The triumph of the Newtonian paradigm initiated a new period of normal science for astronomy (c. 1700—c. 1900). The dominant paradigm consisted of Newtonian mechanics including the law of gravity, and the normal scientist had to use this tool to explain the motions of the heavenly bodies in detail—comets, perturbations of the planets and the moon, etc. In the Einsteinian revolution (c. 1900—c. 1920), however, the Newtonian paradigm was replaced by the special and general theories of relativity.

Further research in the philosophy of science has shown that Kuhn's model, with some modifications, can be extended to mathematics and medicine. Thus Frege's work can be considered as a initiating a revolution in logic analogous to the Copernican revolution in astronomy. The change was from an Aristotelian paradigm, whose core was the theory of the syllogism, to a new paradigm whose core was propositional and first-order predicate calculus.⁴ Then again Semmelweis's investigation can be seen as one of the first steps in a revolution in medicine. The change was from a paradigm whose core was the miasma and contagion theories of disease to a new paradigm with the germ theory of disease as its core.⁵

Now one of the strengths of Kuhn's theory is that it explains why the scientific community made such mistaken judgements regarding figures like Copernicus, Semmelweis and Frege. On Kuhn's model, at the beginning of a revolution almost all the researchers in the field accept the dominant paradigm, and, from the point of view of this paradigm, the new revolutionary approach will indeed seem absurd.

Another important consequence of Kuhn's theory is that the mistaken judgements regarding Copernicus, Semmelweis and Frege are not features of science's past, but are likely to recur over and over again. Of course, long before Kuhn, the Copernican revolution had been studied by historians of science. However, it tended to be regarded as something of a 'one-off' event—a dramatic change which had introduced modern science, but was not likely to recur. This is reflected in the fact that it was often referred to as: *The Scientific Revolution*. Kuhn's originality was to suggest that all branches of science develop through periodic revolutions.

⁴ For more details, see Gillies (1992).

⁵ For more details, see Gillies (2005).

Lessons from the History and Philosophy of Science

This new view was obviously suggested by the revolution in physics in the first few decades of the twentieth century which led to the triumph of relativity theory and quantum mechanics. Kuhn's model of scientific development was roughly as follows. For most of the time we have 'normal science' in which the scientists working in a particular area all, except perhaps for a few dissidents, accept the same dominant paradigm. Within the framework of that paradigm, steady, if perhaps somewhat slow, progress is made. Every so often, however, a period of revolution occurs in which the previously dominant paradigm comes to be criticized by a small number of revolutionary scientists. This small group succeeds in developing a new paradigm, and in persuading their colleagues to accept it. Thus there comes about a revolutionary shift from the old paradigm to a new one. Although revolutions occur only occasionally in the development of a field of science, such revolutions are the exciting times in which really big progress is made in the field.

Kuhn's model of scientific development is, in my view, strongly confirmed by studies in the history of science. Indeed it applies not just to the natural sciences considered by Kuhn, but also to science in the broader sense considered in this paper which includes also mathematics and medicine. In the next section, I will use Kuhn's model to examine the likely effects of the RAE on scientific research in the UK.

7. Analysis of the Likely Effects of the RAE

Let us begin by considering the effects of the RAE on normal science. In a period of normal science, those working in a branch of the subject will all accept the dominant paradigm, and no revolutionary alternative will have been suggested. It will then be an easier matter for the experts in the field to judge who is best according to the criteria of the dominant paradigm. Allocating research funding to these most successful 'puzzle solvers', as Kuhn calls them, will usually enable the normal science activity of puzzle solving to continue successfully. One qualification to this must, however, be introduced on the basis of the discussion of Type I and Type II errors which was given at the end of section 4. We there gave an example of research into a problem, where there are four different approaches to its solution leading to four different research programmes. This situation is still possible, and indeed often occurs, in normal science, for the four different research

Donald Gillies

programmes could all be compatible with the dominant paradigm. As we pointed out, in such circumstances, a thoughtless use of peer review as a tool could easily lead to wrong decisions. Suppose programme 3 in fact turns out to be the one which leads to the solution of the problem, but suppose initially it is supported by only a few researchers. A peer review conducted by a committee chosen at random from those working on the problem might well contain an overwhelming majority of researchers working on programmes 1, 2 & 4, and such a committee could easily recommend the cancellation of funding for research programme 3, a decision which would have disastrous long term results. With this qualification, however, we can say that the RAE is not likely to have too damaging an effect on normal science. The only problem is that normal science tends to be routine in character and to produce small advances rather slowly. Surely, however, we want a research regime to encourage big advances in the subject, exciting innovations, breakthroughs, etc.

It is precisely here that the RAE is likely to fail. Any big advance is likely to have something revolutionary about it, something which challenges accepted ideas and paradigms. However it is precisely in these cases, as we have shown above, that the RAE with its excessive reliance on peer review is likely to have a very negative effect. Our conclusion then is the RAE is likely to shift the UK research community in the direction of producing the routine research of normal science resulting in slow progress and small advances. At the same time it will have the effect of tending to stifle the really good research—the big advances, the exciting innovations, the major breakthroughs. Clearly then the overall effect of the RAE is likely to be very negative as regards research output in the UK.

The RAE is also likely to impact very negatively on the production of wealth-generating science-based technologies in the UK. The reason for this is that the most striking technologies from the point of view of wealth-generation are often based on revolutionary scientific advances. This is well-illustrated by the three examples considered in this paper. Copernicus' new astronomy led, as we have seen, to a much improved navigation, and this was essential to the profitable development of European sea-borne trade in the 17th and 18th centuries. The new mathematical logic introduced by Frege was essential for the development of the computer. It is significant here that Bertrand Russell was one of the first to recognise and develop Frege's work. Russell established an interest in mathematical logic in the UK, which passed on to two later researchers at Cambridge: Max

Lessons from the History and Philosophy of Science

Newman and his student Alan Turing. After the Second World War, Newman and Turing were part of the team at Manchester which produced the Manchester Automatic Digital Machine (MADM). This started running in 1948, and can be considered as the first computer in the modern sense.⁶ Thus Russell's early recognition of Frege's revolutionary innovations led indirectly to the UK taking an early lead in the computer field. This early lead was later lost, as we know, but this was owing to lack of sufficient investment by either the public or private sectors. There was no problem with the UK's research community in those pre-RAE days. Our third case was concerned the revolutionary introduction of antiseptics in conjunction with revolutionary new theories about the causes of disease. We focussed on Semmelweis whose research work was rejected by the medical community of his time. As we remarked, however, Lister was more successful, and was able to persuade the medical community in the UK to accept antiseptics. This was obviously of great benefit to patients, but I would now like to add that it led to very successful business developments. For his new form of surgery Lister needed antiseptic dressings, and he devoted a lot of time and thought to working out the best design and composition of such dressings. As his ideas came to be accepted, the demand for these dressings increased and companies were formed to produce them. One of these was founded by a pharmacist Thomas James Smith. In 1896, he went into partnership with his nephew Horatio Nelson Smith to produce and sell antiseptic dressings. They called the firm Smith and Nephew. Today Smith and Nephew is a transnational company operating in 33 countries and generating sales of £1.25 billion. The company is still involved in wound care as one of its three main specialities, but it has expanded into orthopaedics and endoscopy. One of its well-known products is elastoplast which was developed in 1928. The general design of elastoplast is based on some of the antiseptic dressings developed by Lister. The commercial success of Smith and Nephew is a good illustration of the importance of having a satisfactory research regime in the UK. If Lister's research on antiseptics had met the same fate as that of Semmelweis only 17 years earlier, then the firm of Smith and Nephew would not be with us today.

⁶ For more details about the Manchester Automatic Digital Machine and its claim to be the first computer in the modern sense, see Gillies and Zheng, 2001, 445–9.

Donald Gillies

8. General Conclusions

The RAE is very expensive both in money and in the time which academics in the UK have to devote to it. I have argued in this paper that its likely effect is to shift the UK research community in the direction of producing the routine research of normal science resulting in slow progress and small advances, while tending to stifle the really good research—the big advances, the exciting innovations, the big breakthroughs. Thus a great deal of tax payers' money is being spent on an exercise whose likely effect is to make the research output of the UK worse rather than better. Only one conclusion can be drawn from this, namely that the RAE should be abolished straightaway.

My general argument has brought to light three major faults in the RAE. (1) The RAE rules out the research strategy of working for many years on a piece of research before publication. Yet this strategy has proved very successful in the past. We gave Copernicus and Wittgenstein as examples of the success of this strategy, but many other examples could of course be given. (2) The RAE relies too strongly on peer review, which may work not too badly for normal science, but which can give very erroneous results when it comes to the most important revolutionary advances in science. Frege, Semmelweis and Copernicus were all examples of this. (3) The RAE concentrates too much on trying to eliminate Type II error, that is the error of funding bad research, but devotes no consideration to eliminating Type I error, that is the error of failing to fund good research. Yet Type I errors have much more damaging effects on the progress of research than Type II errors. This was illustrated above all by the case of Semmelweis where a Type I error of failing to recognise and support important research led to thousands of patients dying and a general crisis in the hospitals.

Abolition of the RAE could be accomplished very easily because it would produce no disruption in the system. Indeed research in the UK was very successful for many decades with no RAE. We argued for this in detail in the case of philosophy but the same applies to other areas of research. However, here it might be objected that we can't just go back to the *status quo ante* RAE, for the whole university system and research community has expanded considerably since 1975, and so can no longer be run along the lines which were used in this earlier period. I agree with this point, but would still stress that there is no hurry to introduce a new system for organising research, and that there should be a great deal of

Lessons from the History and Philosophy of Science

thought, discussion and consultation before doing so. What our critique of the RAE has shown is that research is rather a subtle and complicated activity and that producing a regime in which it flourishes is not an easy matter. Perhaps the biggest difficulty lies in the fact that we cannot tell immediately whether a piece of research is good, and sometimes it is only after a period of as long as thirty years that a fairly definite judgement can be reached. In this respect research differs very strikingly from competitive sports such as tennis or football. We can grade tennis players at a particular moment simply by getting them to play each other in tournaments and seeing who beats who. However, we cannot be sure that a researcher whose work is now judged to be of poor quality may not turn out after all to be a Copernicus, a Semmelweis, or a Frege. Even in cases where it is recognised that a scientific discovery has been made, the importance of that discovery may not become apparent for many years. A good example of this is Alexander Fleming's discovery of penicillin which was made in 1928, and published by Fleming in 1929. Fleming was not harshly treated like Semmelweis or Frege, but the significance of his discovery was certainly not recognised immediately. The head of the laboratory where Fleming worked (Sir Almroth Wright) was a Fellow of the Royal Society and a great admirer of Fleming. In 1930 Wright proposed Fleming for the Royal Society citing his discovery of penicillin and some other research achievements. Fleming, however, was not elected in 1930 or in the four following years. In fact Fleming only became a Fellow of the Royal Society in 1943 when he was 62 years old.⁷

But if it is not possible to tell whether a piece of research is good except after a long lapse of time—perhaps as long as thirty years, how can we possibly decide what research to fund at a given moment? It almost looks as if the problem of devising a sensible system for funding research is insoluble. This is not really so, however, and there are ways of overcoming the difficulty. Needless to say, however, they are not along the lines of the RAE. What is needed here is some new thinking and a quite different approach. I have my own ideas of what this different approach might be like, but it would not be appropriate to give them here since my aim in this paper is to give a critique of the existing system. I hope, however, that my paper might be useful to those trying to devise a new system for organising research in the UK by suggesting a way

⁷ These details about Fleming and the Royal Society are taken from Macfarlane, 1984, 140–1, & 202.

Donald Gillies

in which their ideas can be tested. I would suggest that anyone who has thought out a possible research regime should consider a number of major research achievements of the past such as those of Copernicus, Frege, Semmelweis and Wittgenstein, and examine the effect that the proposed research regime would have had on those achievements. If the effect turns out to be negative, then the proposed research regime should be rejected and something better devised to replace it.

References

- Anderson, A.R. (ed.) (1964) *Minds and Machines*, Prentice-Hall.
- Bell, J.L. and Machover, M. (1977) *A Course in Mathematical Logic*, North-Holland.
- Bynum, T.W. (ed.) (1972) *Gottlob Frege. Conceptual Notation and Related Articles*, Oxford University Press.
- Carnap, R. (1963) Intellectual Autobiography. In P.A.Schilpp (ed.), *The Philosophy of Rudolf Carnap*, Library of Living Philosophers, Open Court, 3–84.
- Davis, M. (1988a). Mathematical Logic and the Origin of Modern Computing. In Rolf Herken (ed.), *The Universal Turing Machine. A Half-Century Survey*, Oxford University Press, 149–74.
- Davis, M. (1988b). Influences of Mathematical Logic on Computer Science. In Rolf Herken (ed.), *The Universal Turing Machine. A Half-Century Survey*, Oxford University Press, 315–26.
- Edmonds, D. and Eidinow, J. (2001) *Wittgenstein's Poker. The Story of a Ten-Minute Argument between two Great Philosophers*. faber and faber.
- Frege, G. (1879) *Begriffsschrift, eine der Arithmetischen nachgebildete Formelsprache des reinen Denkens*. English translation in Bynum (1972), 101–203.
- Gillies, D.A. (1992) The Fregean Revolution in Logic. In D.A.Gillies (ed.) *Revolutions in Mathematics*, Oxford University Press, 265–305.
- Gillies, D.A. (2002) Logicism and the Development of Computer Science. In Antonis C.Kakas and Fariba Sadri (eds.) *Computational Logic: Logic Programming and Beyond*, Part II, Springer, 588–604.
- Gillies, D.A. (2005) Hempelian and Kuhnian approaches in the Philosophy of Medicine: the Semmelweis case, *Studies in History and Philosophy of Biological and Biomedical Sciences*, **36**, 159–181.
- Gillies, D.A. and Zheng, Y. (2001) Dynamic Interactions with the Philosophy of Mathematics, *Theoria*, **16**, 437–59.
- Hempel, C.G. (1966) *Philosophy of Natural Science*, Prentice-Hall.
- Kuhn, T.S. (1957) *The Copernican Revolution*, Vintage, 1959.
- Kuhn, T.S. (1962) *The Structure of Scientific Revolutions*, University of Chicago Press, 1969.

Lessons from the History and Philosophy of Science

- Lucas, J.R. (1961) Minds, Machines and Gödel, *Philosophy*, **36**. Reprinted in Anderson, 1964, 43–59.
- Macfarlane, G. (1984) *Alexander Fleming. The Man and the Myth*, Chatto & Windus
- Malcolm, N. (1958) *Ludwig Wittgenstein. A Memoir*, 2nd Edition, Oxford University Press, 1989.
- Mendelson, E. (1964) *Introduction to Mathematical Logic*, Van Nostrand.
- Monk, R. (1990) *Ludwig Wittgenstein. The Duty of Genius*, Jonathan Cape.
- Pitcher, G. (1964) *The Philosophy of Wittgenstein*, Prentice-Hall.
- Semmelweis, I. (1861) *The Etiology, Concept, and Prophylaxis of Childbed Fever*. English Translation by K. Codell Carter, The University of Wisconsin Press, 1983.
- Turing, A.M. (1950) Computing Machinery and Intelligence, *Mind*, **59**. Reprinted in Anderson, 1964, 4–30.
- Wittgenstein, L. (1953) *Philosophical Investigations*. English Translation by G.E.M. Anscombe, Blackwell, 1963.