Study of Big Science Groups Hits Raw Nerve

Some members of Britain's scientific establishment are upset by a study comparing the quality of research at major scientific facilities

Paris. An attempt by two British researchers to determine the relative quality and productivity of research groups at major scientific facilities has stirred up a bitter debate in Britain. One leading British scientist even threatened to take legal action to block publication of some of the researchers' conclusions. In particular, much of Britain's scientific establishment is upset by the researchers' harsh criticism of the way decisions on "big science" projects have been made in the past.

The two researchers, Benjamin R. Martin and John Irvine of the University of Sussex's Science Policy Research Unit (SPRU), have developed a method they describe as "converging partial indicators" to measure the relative productivity of matched research teams at major facilities. The results of their study are published in the current issue of *Minerva*.

Their conclusions on scientific output contain few surprises. In radio astronomy, for example, they show that in the 1970's Britain's large-scale interferometer at Cambridge was producing more scientific papers of greater individual impact than the radio telescope at Jodrell Bank near Manchester. In high-energy physics, Stanford's 20-GeV linear accelerator (SLAC) comes out top of the league in the same period, according to various different measures, with Britain's 4.5 GeV electron accelerator (NINA) at the Science Research Council's Daresbury Laboratory well down the list.

More controversial are the implications they draw for the organization and control of British science. They suggest that where mistakes have been made, these have been the result of an excessive concentration of power and of the influence of particular groups in the scientific community.

The *Minerva* paper concentrates on Britain's experience in high-energy physics. Together with two other studies—one looking at optical astronomy and the other at radio astronomy—it was produced as the result of a research project sponsored by Britain's Social Science Research Council into the funding of major facilities. One goal of the work has been to improve the quality of scientific decision-making in a period of budgetary austerity.

Martin and Irvine start from the premise that many of the criteria conventionally used to measure scientific productivity, such as the total output of scientific papers and other forms of citation analysis, contain potential flaws that limit their usefulness. However, they suggest that, if several such criteria are used, and if the conclusions all point in the same direction, then they can be taken with much greater confidence as a valid assessment of the relative performance of different research groups. Hence the name "converging partial indicators."

Three measures of scientific quality have been used. The first is an assessment of the significance of the total number of papers produced by different scientific teams working at a particular facility. The second tries to quantify the scientific impact of the work carried out at the facility by looking at the frequency with which the papers produced there are cited by other scientists in the field. Thirdly, the two researchers carried out interviews with scientists at each of the facilities they studied, as well as other major scientific centers, eliciting from each a subjective assessment of the work at their own facility and at rival sites.

> "The scientific community seems to feel threatened by external review of this type," says Irvine.

In each of the fields looked at, claim Martin and Irvine, the results reached by these three separate—though, they admit, not entirely independent—types of measurement point toward the same conclusions in terms of relative scientific productivity.

The current *Minerva* paper, for example, shows that, up to the time it was closed in 1977, NINA produced an average of six papers a year compared with eight for the Cambridge Electron Accelerator (CEA) in Cambridge, Massachusetts, and 13 for the Deutsches Elektronen Synchroton (DESY) in Hamburg, West Germany. Comparing the scientific impact of the accelerators, they claim that NINA failed to produce any paper cited more than 19 times a year, while CEA produced 10, DESY 15, and the more powerful SLAC machine 56. Finally, a subjective ranking by 80 European and U.S. physicists placed NINA below most western accelerators and ahead only of Soviet competitors.

A similar consistency exists in evaluations of the 2.5-meter Isaac Newton optical telescope (INT), which was recently moved from southern England to the Canary Islands. Before its removal, the INT fared badly on all three criteria compared to three similar facilities, the 3-meter Lick Telescope on Mount Hamilton in California, the 2.1-meter telescope at the Kitt Peak National Observatory in Arizona, and the 1.5-meter telescope at the Cerro Tololo Inter-American Observatory in Chile.

Martin and Irvine have used the same techniques to compare four radio-astronomy observatories: the radio telescopes of the Nuffield Radio Astronomy Laboratories at Jodrell Bank and the Max-Planck Institute for Radioastronomy in Bonn, West Germany; and the interferometers at the Mullard Radio Astronomy Observatory at Cambridge, England, and the Netherlands Foundation for Radio Astronomy at Westerbork. According to an early version of their results, Cambridge comes out top of the four in terms of scientific output over the period 1969 to 1978 and Jodrell Bank at the bottom.

None of these assessments has caused much surprise among those familiar with the work of the different centers. Tom Kibble, for example, head of the department of physics at Imperial College in London, claims the results "tend to reinforce the prejudices that I already have." He points in particular to the consistency of judgment implied by the converging views of physicists in different countries about the relative merits of major accelerators. "I would have expected rather bigger differences," he says.

Irvine and Martin have gone beyond their empirical analysis, however, in various directions, and this is where the controversy has started. Firstly, they suggest that their method, unlike some previous attempts to measure the quality of scientific outputs, "seems to be able to produce in a systematic, rigorous and reproducible manner, data that the scien-

tific communities concerned generally accept as providing an accurate reflection of the past performance of different research groups." Such data, they suggest, should be used more frequently in determining the future allocation of research funds. This conclusion is supported by physicist John Ziman, who argues that the study makes a good case for using convergent indicators to evaluate past research performance.

The two Sussex researchers also use their analysis to make judgments about past decisions by British scientists and research administrators. In the case of NINA, for example, which was closed down in 1977 despite protests from a number of physicists, Irvine claims that, if a rigorous assessment of the accelerator's poor scientific record had been available, "It would have been closed much earlier."

They also criticize the original decision to locate the INT in Sussex rather than abroad. Britain's poor weather and local viewing conditions combined to place severe limits on the observations that could be made from it. However, part of the reason for the low productivity, they suggest, was the way the telescope was run by the Royal Greenwich Observatory (RGO). They repeat astronomers' complaints of a "civil service mentality," of a failure to provide adequate backup to university scientists, and of a low level of instrument innovation, suggesting that each could have contributed to the low productivity. And they quote the conclusion of a study that the new La Palma Observatory should be run by a body other than the RGO, a recommendation which was not implemented.

Both their methods and their conclusions have generated sharp responses. Graham Smith, the current astronomer royal, who was responsible as director of the Royal Greenwich Observatory for establishing the INT in its new home on the Canary Islands, disagrees that the method of "partial indicators" is necessarily valid. "To say that, if you have three different methods which seem to agree on a conclusion, then the conclusion is sound, is nonsense."

Charges of mismanagement of the INT in the past are similarly sharply disputed by Smith, who is now director of Jodrell Bank. "No one is disputing the figures about the productivity of the telescope," he says, pointing out that the decision to move the telescope was taken by a consensus of British astronomers. "But the difference is almost entirely due to the geographical conditions; the whole point is that it was not a good site."

Smith's predecessor at Jodrell Bank, Sir Bernard Lovell, has been even more vociferous in his criticism of the Sussex group's conclusions about the performance of the radio telescope under his direction. Lovell threatened to take legal action against the journal *Research Policy* if it published an early version of their results. According to a summary in the French magazine *La Recherche*, which appeared in 1981, Martin and Irvine calculated that not only did Jodrell Bank produce only one paper between 1969 and 1978 cited more than 15 times a year (compared to 12 from the Cambridge interferometer) but the costs per article published were almost twice as high as those for papers from the comparable German radio telescope.

Stanford's linear accelerator

Top of the league in the 1970's

Lovell and Smith both claim that the costs have been distorted. So too, they say, has the radio telescope's contribution to science, for example by ignoring the groundwork that it has provided for other radio astronomy projects. The legal threat was lifted after the costs had been recalculated by the Sussex researchers and after some of Lovell's arguments were included in the paper. Publication of the revised paper is now expected "in the near future," according to SPRU director Geoffrey Oldham.

Given their controversial nature, the Sussex results have received substantial coverage in the British press. Irvine and Martin have done little to dampen the flames. A widely-distributed press release on their work described the INT as "a failure" with a "startlingly poor record" of "third rate research" that had been "ill-starred" from the outset. When reporters started to follow up the story, the press release was hastily withdrawn. Its substitute talked more modestly of a "rather poor record" of "mediocre research." But it was the language of the first release that still found its way into many popular newspapers.

Unsurprisingly, the scientific community has not taken kindly to this type of treatment. "The sort of facts that lie behind policy decisions should be studied more widely, but this way of doing it is doing a lot of damage," says Smith. "It seems that these people are biased against establishment science."

Irvine admits that "we may have gone a bit far in trying to explain why some centers did well and others did badly." But he suggests that the criticism of their work is partly motivated by self-defense, and that public critics are often more muted in private. "These people have to defend their interests," he says, suggesting that "the scientific community seems to feel threatened by external review of this type."

What many science administrators see as the make-or-break research project for the Sussex pair is a major study currently being completed of decision-making at the European Organization for Nuclear Research (CERN) in Geneva. CERN, which is supported by contributions from 12 member states, came out of the earlier studies relatively well. Although it placed second behind SLAC in the period 1969 to 1978 and only slightly ahead of other U.S. accelerators, CERN director Herwig Schopper suggests that this is partly the result of the dates chosen for analysis and partly the result of luck. Since 1978, says Irvine, all the evidence points to a significant weakening in the previous U.S. dominance of the field.

Due to their potential political sensitivity, the results of the CERN study are being carefully checked by laboratory staff before they are published. While not casting doubt on the decision to proceed with the current construction of the Large Electron-Positron storage ring (LEP), the study is said to question the way in which some of the early planned experiments on LEP have been selected, as well as some of the decisions taken at CERN in the 1970's.

Research administrators in Britain are waiting to see the results of the CERN study before concluding whether or not Irvine and Martin's method is as useful as they claim in allocating resources to new facilities. Physicists are also remaining wary. "By and large their conclusions seem to be reasonable, but many physicists are a bit suspicious of the whole approach," says Kibble at Imperial College. "Partly it is a question of self-interest," he admits. "People do not like their dirty linen washed in public."

—DAVID DICKSON