3. See, for example, J. Irvine and B. Martin, 'Assessing Basic Research: The Case of the Isaac Newton Telescope', *Social Studies of Science*, Vol. 13 (1983), 49–86; Martin and Irvine, 'An Evaluation of the Research Performance of Electron High-Energy Physics Accelerators', *Minerva*, Vol. 14 (1981), 408–32; Martin and Irvine, 'Assessing Basic Research: Some Partial Indicators of Scientific Progress in Radio Astronomy', *Research Policy*, Vol. 12 (1983), 61–90.

4. A general criticism of 'methodological triangulation' is that it is not clear what one does when the measures do not converge. I agree with Krige and Pestre that non-convergence is the more interesting and demanding case. As will be seen below, I do not agree that IM's measures converge on scientific quality in any straightforward way, since priority determines rewards to too great an extent. See J. Krige and D. Pestre, 'A Critique of Irvine and Martin's Methodology for Evaluating Big Science', *Social Studies of Science*, Vol. 15 (1985), 525–39.

5. P. Medawar, *The Art of the Soluble* (London: Penguin, 1969), 96.

6. Some economists would say that scientific rewards are a 'positional good'.

7. See H. M. Collins, *Changing Order: Replication and Induction in Scientific Practice* (London and Beverly Hills, Calif.: Sage, 1985).

8. See J. Irvine, I. Miles and J. Evans, *Demystifying Social Statistics* (London: Pluto Press, 1979).

Author's address: Science Studies Centre, School of Humanities and Social Sciences, University of Bath, Claverton Down, Bath BA2 7AY, UK.

Responses and Replies (continued)

Evaluating the Evaluators: A Reply to Our Critics

Ben R. Martin and John Irvine[1]

We should begin by welcoming the opportunity to have our work critically reviewed in this journal. Given the wide range of differences we have with our critics, it is perhaps first worth identifying the issues on which we are in agreement. Four stand out in particular:

Social Studies of Science (SAGE, London, Beverly Hills and New Delhi), Vol. 15 (1985), 558–75

(1) research addressing science policy issues is urgently needed;

(2) advances in the area of science studies over the last two decades mean that it is now possible to tackle substantive science policy issues in a more systematic manner;

(3) bibliometric studies have begun to focus on 'issues of importance' and 'have made a marked impact',[2] even succeeding in certain cases in providing 'a wealth of statistical data to be taken into account by future policy makers';[3]

(4) our approach is certainly capable of being improved upon.

In view of the unanimity on these points, we are disappointed at how little our critics have to offer by way of concrete suggestions as to how one might improve the techniques we have been attempting to develop to provide systematic and reliable information for science policymaking. Let us consider in turn the points made in the four critiques and the improvements (if any) in evaluation methods and approaches that they suggest.

Policy Research or Fairy Tales?

Harry Collins begins by congratulating us for provoking 'some serious questioning of British science policy'. 'Nevertheless', he continues, 'their research is not without flaws, nor is it obviously the best way for science policy to make use of recent work in science studies'.[4] What are these 'flaws'?

The first is that we have taken as our unit of analysis research institutions rather than units reflecting cognitive boundaries.[5] According to Collins, policy research should 'start by disaggregating science according to cognitive rather than institutional boundaries'.[6] This is a fundamental difference, and one on which we disagree strongly with Collins. In basic science, at least, policymakers often take decisions on whether to fund research groups, departments, laboratories, and centres, as opposed to entities defined purely by cognitive boundaries. While the latter may be the most suitable unit of analysis for the study of certain questions in the sociology of science, for science policy research the former is important and clearly cannot be ignored. Examination of the expenditure of most national research-funding agencies shows that research groups, laboratories, and so on — that is, institutionally defined entities — account for a large proportion of their total budgets (some grants are, of course, made to individuals — most importantly to open up new areas — but these account for only a fraction of the funds). There are, of course, other cases, generally involving more applied research, where such organizations may have to decide whether or not to support a field (and in our more recent work we have tried to address this type of question[7]). However, even then, once an agency has decided to support a field for strategic or other reasons, it still has to choose which research groups or institutions to support, so information relating to their past and likely future performance is just as vital. In short, Collins is wrong to assume that what is most interesting from the point of view of a particular sociology of science standpoint coincides exactly with the main interests of policy-makers. To paraphrase Collins, 'Science policy should be about maintaining efficient institutions *as well as* maintaining the sort of cognitive community that will produce the desired scientific products'. (As an aside, we would also point out that much interesting sociology of science would be lost if Collins' demarcation were to be universally adopted.)

The second criticism is that the indicators we have employed reflect the atomistic reward system of science, with all the prizes going to those who are first. In fact, only one of our indicators (highly-cited papers) relates directly to 'discoveries' or 'coming first' — the others relate much more to taking 'part in a scientific endeavour'.[8] For example, our study of world high-energy physics showed that between 1961 and 1982 there were very few occasions on which CERN was 'first' (nearly all the crucial discoveries and advances over this period were made in the United States).[9] However, the other indicators suggest that CERN accelerators were perhaps the most successful in terms of experiments yielding more precise measurements and better statistics, a finding in line with the results obtained from 180 interviews.[10]

Collins then attempts to demonstrate by means of 'a fairy tale' 'the difference between foreseeable policies based on the reward system and institutional units, and those based on cognitive criteria'.[11] The tale, we are informed, 'rests on some substantial, but not totally unrealistic, assumptions about how science works'. One assumption (apparently unrecognized by Collins) is that the international dimension of science can be completely ignored. (This assumption may actually hold in Collins's own research area, but that is something of an exception.) Besides Paragon and Querulous in country (or continent) A, there would be other Paragons and Querulouses around the world. Hence, if a British Research

Council, for example, were to decide to close the world leader in a particular field, the probable result would not be that the same discoveries would be made six months later by a less costly UK group, but by Texas Megabuck Lab or Nippon Lab. Furthermore, Collins's fairy tale, even if it had been based on realistic assumptions, entirely misses the point. In most cases, those *outside* the field concerned will not know that Paragon has been making ten major advances a year while Querulous has been making none. Scientists within the field will undoubtedly have subjective impressions of the difference between the two, but, for reasons related partly to the existence of vested interest groups, they may not reveal the true extent of the difference to others. Nor are they likely to have as wide-ranging an understanding of the factors structuring research performance as that, for example, yielded in our study of high-energy physics accelerators. We have seen our task as one of providing systematic information on research performance and the factors structuring it, in a form accessible not just to researchers in the field concerned but also to policymakers, scientists in other fields, politicians and the public, thereby making possible greater transparency in the scientific decision making process.

Having argued that our work is not 'the best way for science policy to make use of recent work in science studies'[12] (and employed his fairy tale to prove that 'different ways of approaching the measurement of scientific quality can lead to widely different conclusions'[13]), what does Collins suggest might constitute 'the best way' forward? As his first example, he proposes that 'we could begin to investigate the balance of the different "phases of science" (normal and extraordinary) and the conditions of their survival'.[14] But does this (or indeed any of the other examples Collins cites) actually correspond to the primary concerns of policymakers? Over the last seven years, we have devoted considerable effort to trying to establish just what are their main interests and problems. Almost all the many policymakers in the UK and overseas who have discussed this with us have stressed the need for better information on the performance of research groups, departments, facilities and laboratories, and on where their country stands in international terms in particular fields.[15] As far as we can recollect, not one has mentioned any of the examples listed by Collins.[16]

The Case of the Misread Papers

It is difficult to know how to take the criticisms by Robert Bud when there appears to have been such a fundamental misreading of our work as to lead him to categorize high-energy physics as 'strategic research'. In our book, *Foresight in Science*, we explain why the conventional classification of R&D (enshrined in the OECD 'Frascati Manual') into 'basic research', 'applied research' and 'experimental development' is increasingly inadequate. In particular, within basic research one can now identify two very different types of research activity, namely 'curiosity-oriented research' and 'strategic research' — the latter being work where no specific end-product or process can yet be identified (so it is not 'applied' research) but where the research is expected to produce a broad base of knowledge likely to form the background to the solution of practical problems. The book then concentrates on methods for looking at the longer-term future for strategic research. It does not, as Bud claims, dismiss curiosity-oriented research as an empirically residual class, but rather argues that its greater intrinsic unpredictability makes attempts at longer-term forecasts less reliable and therefore not so worthwhile. As with any categorization, 'curiosity-oriented research' and 'strategic research' are to some extent ideal types, but it is not difficult to think of examples of research activities which are predominantly one or the other. Most basic research biotechnology, for example, is primarily 'strategic', while at the other end of the spectrum we would place almost all astronomy and high-energy physics. We can only assume that Bud's view of high-energy physics as strategic research is linked to his belief (stated in two places) that experimental high-energy physics is carried out on 'reactors'[17] and therefore is presumably in some way part of nuclear energy research. In fact, the two fields parted company in the 1950s, and the links between them (both cognitive and social) are now almost non-existent (or at least no greater than with other fields of physical science). This is the reason why in our study we concentrated on evaluating the performance of accelerators in scientific terms. We would, however, welcome positive suggestions on how to extend the evaluation to cover 'extrinsic' goals.[18]

Putting aside the doubts engendered by such a mistake, what can we make of Bud's criticisms? We are told that we have devoted insufficient attention to examining the goals of scientific research (although Bud does not actually tell us what he thinks these are), and that we have used indicators 'which do not indicate levels of achievement of those hypothetical [sic] goals'.[19] Right from the start of our research programme, we have explicitly recognized that one cannot use the same indicators to evaluate different types of research. If Bud were to read, for example, the report on a study of mechanical, electrical and electronics engineering that we carried out for a 1981 Norwegian Royal Commission, he would see that a very different approach and set of indicators was adopted to evaluate these more applied areas, whose objectives clearly differ from those, say, in high-energy physics. (We laid great stress, for example, on 'customer review'—that is, obtaining the views of industrial firms on the research activities being assessed—and produced a large volume of quantitative data relating to this.)[20]

A second criticism relates to our comparison of the research performance of accelerators in the Eastern bloc, the United States and Western Europe, with Bud pointing out that the accelerators 'are embedded in three such different cultural systems'.[21] Whereas one response to these cultural differences may be to throw up one's arms in horror and imply that no valid comparisons can apparently ever be drawn, ours is to attempt to identify variations in scientific output and impact and then examine the extent to which these can be related to the cultural differences. (Eastern and Western high-energy physicists are not in fact the two completely distinct and non-interacting communities that Bud implies. They attend many of the same conferences, use each other's accelerators to a limited extent, and publish in essentially the same body of international journals, with East Europeans having made increasing use of West European journals over the last twenty years.)

Next, Bud chides us that 'the interpretation of the quantitative data is perfunctory. Conclusions are drawn only about overall magnitudes'.[22] As with any piece of research (but particularly one in a new area), the analysis could have been taken further. We have not attempted to solve all the problems in making East-West comparisons of scientific performance at once and have also been careful to avoid drawing over-ambitious conclusions (that is to say, the caveat has not 'disappeared' as Bud claims) which cannot be adequately supported by data with all the limitations we describe. If Bud has specific suggestions to make in relation to interpreting the empirical data further, we would appreciate the opportunity to consider them.

Overall, Bud, like Collins, seems to exhibit a myopic tendency to see everything from the perspective of science studies rather than science policy. His 'problem with all the indicators is the same. What are their implications for our understanding of the relationship between the scientific communities?'[23] We would merely note that this is not a problem that science policymakers have identified to us as one of their principal concerns.

The CERN Critique

The stated aim of the critique by John Krige and Dominique Pestre is simple and admirable — it is to defend 'intellectual rigour'.[24] It is therefore alarming to find them beginning the presentation of their case by quoting what someone stated at a seminar had been said to him on some other unspecified occasion by 'a Chinese colleague' who was in turn 'quoting' two other people. (Did Marx and Engels actually both say anything as simplistic as that — we would be interested in seeing the relevant references!) Yet, later in their Response, this unsubstantiated anecdote has been elevated to the status of evidence which 'makes the point'. Is this more 'rigorous' than our structured interviews with over 180 high-energy physicists, each of whom then completed a detailed attitude survey?

After presenting a clear summary of our work on big science, Krige and Pestre commence their critique by arguing that such non-scientific benefits as 'manpower' training, technological spin-off and national prestige are not 'secondary reasons' for supporting high-energy physics. This would seem to suggest that they see such benefits as 'primary' — that is, of equal (or greater?) importance than the scientific question of how much the research will increase our knowledge of the material world. (Krige and Pestre do actually contradict this two pages later: 'We are not claiming that "non-scientific" considerations, rather than the production of scientific knowledge, are "primary" or "more important" in assessing Big Science.'[25] But if they are neither 'primary' nor 'secondary', what exactly are they? A little more rigour is required here.) This raises the interesting (and unanswered) question, 'How poor would the scientific performance of CERN have to become before scientific criteria finally over-rode non-scientific ones?' Or are we to agree with Moed and van Raan that centres like CERN 'are simply too rare, too prestigious' ever to close down?[26]

In concluding our analysis of CERN's past performance, we discussed the political, educational and technological benefits it has yielded.[27] We are, however, less convinced than our CERN colleagues by the findings of Schmied on technological spin-off from CERN.[28] While it may be comforting to CERN to have a study claiming to find that the industrial contracts placed by CERN have resulted in increased 'economic utility' several times greater than the original value of those contracts, this should not blind us to *its* methodological inadequacy. First of all, the concept of 'economic utility' employed in the study would be unrecognizable to most economists. Secondly, the methodological approach of asking firms to identify subsequent contracts that have drawn upon work originally carried out for CERN and then counting their total value as an economic benefit from CERN is extremely dubious. Thirdly, there is the question of how much reliability one can place on the answers of firms dependent to a greater or lesser extent on CERN for future business when questioned in a study financed by CERN. Lastly, the opportunity costs are completely ignored — 'is the level of technological spin-off higher than it would have been if the resources spent on CERN had instead been used to support some other type of research, such as exploration of the ocean bed, for example?'[29] Given these various doubts, we would place rather less confidence in the results of this study than do Krige and Pestre.

While we recognize the importance of the various educational, technological and political benefits associated with CERN, and would have devoted more effort to evaluating CERN in terms of them if time and resources had permitted (we certainly do not dismiss them as 'trivial' as Bud claims[30]), we would still argue that CERN in particular, and Big Sciences like astronomy and high-energy physics in general, are supported more for the contributions they are expected to make to scientific knowledge.[31] In our papers on CERN,[32] we quote various explicit statements from CERN to this effect.[33] Furthermore, few of the scientists involved attempt to defend the funding for their work by reference to 'extrinsic' benefits: our attitude survey revealed that twice as many felt that expenditure on the field could be justified only in scientific terms as believed it could in terms of the spin-off it generates. Krige and Pestre's observation that Big Sciences like high-energy physics 'have a kind of role analogous to that of military research in the economical field' is perhaps more revealing than they intend, given the extensive literature demonstrating how military R&D, while it does yield occasional important spin-offs, at the same time locks up scarce R&D resources that might more profitably be employed in other forms of research.[34]

The next point made by Krige and Pestre is that we 'rely essentially on the *average feelings* of the scientific community, leading us to doubt whether they can produce assessments substantially different from those obtained through a more classical panel system'.[35] Here, it is important to distinguish between conventional peer-review (involving a small number of referees or 'experts' on a panel) and our extensive peer-evaluations drawing in very large numbers of researchers across different countries and based on structured confidential interviews and attitude surveys. Because the latter approach yields relatively consistent results, it does not logically follow that conventional peer-review is 'rather reliable'.[36] Indeed, we have encountered several instances where the existence of an 'oligopolistic' situation in a research field has led to the two approaches yielding very different results. One of the most prominent (and costly) involved the ISABELLE project in the United States,[37] where conventional peer-review continued to suggest that this was the top priority for US high-energy physics long after most researchers had privately recognized that this was no longer true (and admitted as much to us in interviews). By the time the project was finally aborted, some $200m had been spent.

Another important problem with conventional peer-review is the inherent lack of accountability to those outside the specialty concerned. While it may be easy for a panel of high-energy physicists, for example, to conclude that existing accelerators are successful and that a proposed new facility is an absolute priority, such statements can equally easily be dismissed by outsiders on the 'Rice-Davies' grounds that 'They would say that, wouldn't they?' Hence the need to complement traditional peer-review with bibliometric data on research performance, or external assessments of the likely future performance of new facilities. If presented in a form accessible to those outside the field, such information can play a major role in providing evidence to other scientists competing for scarce funds (as well as to policymakers and the general public) that decisions are well founded. In this way, one can begin to achieve more open and transparent decision-making in basic science.

Krige and Pestre, in fact, provide a good example where such accountability may have been somewhat lacking. At the time it was built, the CERN SPS accelerator was criticized by some, particularly in the United States, for being over-engineered and 'gold-plated' (it cost almost twice as much as the similar-energy Fermilab accelerator). Yet, as Krige and Pestre point out, it was this which permitted the SPS to be converted into a proton-antiproton collider more quickly and at lower cost than the Fermilab machine, and hence to discover the Intermediate Vector Bosons. However, this raises several awkward questions: how many other instances of 'gold-plating' have there been which have not paid off in this way? At what cost? And were CERN member states aware, when they agreed to the SPS project, that they were funding a comparatively expensive machine which might or might not turn out to have advantages over a cut-price version? As defenders of intellectual rigour, our critics must recognize that one example where 'gold-plating' did pay off by no means proves that this policy always constitutes the best use of limited resources.

As for the 'rhetorical rules' identified by Krige and Pestre, we take these as not wholly uncomplimentary. We would, for example, prefer to be accused of appealing to common sense than the alternative, whatever that might be. However, on a point of information in relation to the first rule, we do not '*envisage* all possible objections first',[38] in the sense of dreaming them up ourselves beforehand. At the end of the CERN study, we spent a year circulating drafts of the papers for comment to around 100 high-energy physicists, other scientists and officials in funding agencies, then revising and recirculating them. We also debated the results with the scientific community in various seminars. This was the way in which possible objections were identified, and then taken into account before submitting the papers for publication. It is our view that work on science policy must be capable of being defended in front of a scientific audience. Given the emphasis that we have placed on such validation of our results by the scientific community, we would dispute the claim by Moed and van Raan that there is no test of the 'validity' of our method.[39]

The Leiden Alternative?

This brings us to the criticisms of Henk Moed and Anthony van Raan. The first relates to the various indicators that we have employed and the extent to which they converge. The indicators, although related to some extent, do nevertheless reflect slightly different facets of research performance. Publication totals give an indication of the overall scientific production of a research group, while numbers of papers per researcher or per dollar reveal something about its productivity. The average number of citations per paper is a measure of the impact those publications have on the scientific community, while peer-rankings (where scientists rank the contributions of different institutions) provide evidence on the perceived significance of the results. Lastly, data on the distribution of highly-cited papers reveal which groups have been responsible for the few key 'discoveries' or advances in a specialty, while aggregate citation statistics reflect the overall impact of a large number of incremental additions to knowledge. (Contrary to what some critics suggest, we do not claim that any of the indicators measure the 'quality' of research.[40]) As we have spelt out in detail,[41] all these indicators are 'partial' in nature. However, in the real world of science policy, one has to accept that there are no perfect measures of scientific performance and use whatever indicators are available, recognizing their limitations and working with them as best as possible.

The starting point when we began work on research evaluation was the hypothesis that, if the various indicators were applied to matched research groups, we would expect them to yield broadly convergent results. This they did in the case of electron high-energy physics, radio astronomy and optical astronomy. In the latter, for example, we obtained the results summarized in Table 1. We are surprised by Moed and van Raan's claim[42] that the indicators do not converge in this instance. As we stated when reporting the very first research evaluation results, the fact that the indicators converge in a given case does not 'prove' that the results are 100 percent certain — the indicators may all be 'wrong' together.[44] However, if a research facility like the Lick 3-metre telescope produces a comparatively large publication output at fairly low cost, if those papers are relatively highly cited, if it yields many of the highly-cited papers in the field, and if large numbers of astronomers rate it highly in the course of structured interviews, we would place *more* credibility on the resulting conclusion that this was a successful facility than if the same finding was arrived at by a panel of three or four 'experts' without access to the sort of systematic information that we have collected. Similarly, if a telescope like the INT produces relatively few papers at high cost which receive comparatively few citations in total (even though each paper on average has a reasonable citation-

TABLE 1
Output Indicators for Optical Telescopes — A Summary[43]

	Lick 3-metre	KPNO 2.1-metre	CTIO 1.5-metre	INT 2.5-metre
Average no. of papers p.a., 1969–78	42	43	35	7
Cost per paper in 1978	£13k	£7k	£6k	£63k
Citations to work of past 4 years in 1978	920	710	580	140
Av. citations per paper in 1978	4.2	3.3	3.3	3.6
No. of papers cited 12 or more times in a year, 1969–78	41	31	21	4

per-paper figure), and if it yields rather few highly-cited papers and is ranked towards the bottom of a list of 12 telescopes by 50 astronomers, we would be reasonably confident that its scientific performance had not been particularly good in world terms. Even so, we have always stressed that such findings need to be interpreted with care, and that, rather than being used to replace the peer-review process, they should be fed into it to enhance its effectiveness and help keep decision-making 'honest'.

As for the concept of 'convergence', like any notion when first formulated it may initially have been somewhat 'poorly developed'. However, as more empirical studies have been completed, so the conditions under which the indicators might be expected to converge have become clearer. First, the less satisfactorily research groups are 'matched', the less convergence is to be expected. The convergence for the four radio astronomy centres was reasonably good — certainly much better than when we compared CERN (which then had three major machines) with other high-energy physics laboratories (mostly operating just one accelerator each). Secondly, in a period of revolutionary change within a field, the facility producing the most highly-cited papers and 'discoveries' (as SLAC did in experimental high-energy physics during the mid-1970s) is likely to be judged the most successful even if it is not the world leader in terms of numbers of publications and citations. Thirdly, convergence may be expected if the indicators are applied to research groups working in an *internationally homogeneous* field, but not if the field consists instead of several distinct, non-interacting communities (which do not attend the same conferences, publish in the same journals, cite each other's work, and so on).[45] For situations between these extremes, where there is some interaction but not complete international homogeneity (as we encountered when comparing high-energy physics in the East and West[46]), one may, for example, first have to adjust the indicators for differences in publication and referencing practices. This is, however, an area where more work is needed.

Another disagreement we have with our Dutch critics concerns the respective merits of computerized and manual-scanning approaches to research evaluation. We have no objection to computing *per se*, and indeed use outside data-bases where there are significant advantages — for example, the *NSF Science Literature Indicators Data-Base*, which we have used to draw conclusions about the overall scientific performance of countries across research fields.[47] However, there are certain tasks in research evaluation where a manual scanning approach is the only option:

(1) In some cases, computerized scanning is too expensive because of the high access costs to data-bases (research evaluations should not cost more than a fraction of the research being evaluated).[48]
(2) There can be difficulties with using existing computerized data-bases in defining the boundaries of fields.[49] In the case of the CERN study, for example, it was necessary to construct our own data-base on experimental high-energy physics. Here, there was no option to a manual-scanning approach because of the need to read large numbers of physics papers in order to establish which related to experimental high-energy physics and which did not.
(3) Without reading the papers, it is generally impossible to establish which research facility has been used to produce the experimental results (this is necessary for carrying out analyses of institutional and/or national performance).
(4) Similarly, reading is essential to distinguish between experimental and theoretical papers in a given field, something which is vital from a policy point of view for areas where the costs of experimental and theoretical work are very different.

It is therefore misleading to pretend that the choice between manual and computerized scanning 'all depends on the search-algorithm'.[50]

For the tasks mentioned above, the best search-software is still to be found in the human brain. Furthermore, the worry about 'the scanner getting tired (or even crazy)'[51] needs to be counterbalanced by the benefits associated with actually reading the papers and becoming immersed in their content.

Since Moed and van Raan hold up their own work as an exemplar of bibliometric assessment, it should be pointed out that their approach is in our view somewhat flawed in that performance indicators are applied to different departments within a single university — that is, they are used to draw comparisons between fields characterized by very different publication and citation practices.[52] There is no attempt to compare 'like' with 'like'. While it may be relatively simple for those with sufficient funds to obtain bibliometric data for such groups (by ordering computer print-outs from commercial data-banks), we harbour grave misgivings about applying indicators in such a fashion, and would be interested to know what the researchers thus assessed think of the validity of the results.

Moed and van Raan's final criticism is that 'performance analyses of large facilities such as accelerators have a limited relevance for research policy'.[53] In their view, 'as soon as, at least in the West, a specific facility — like an accelerator — loses its position at "the front", scientists will move to other facilities. . . . It would . . . be [more] relevant to policy to follow specific (groups of) scientists . . . These groups . . . are, in our opinion, the most interesting (and therefore most policy-relevant) "level" to evaluate.'[54] This criticism is misplaced for two reasons. First, it assumes a perfect labour market within science completely at odds with the actual situation (even in the West).[55] Secondly, central facilities account for a large proportion of the expenditure by national research-funding agencies,[56] and for Big Sciences like high-energy physics the main policy decisions focus on whether to fund a centre or a new accelerator or detector, not on which university user groups to support.

Concluding Note

Let us conclude as we began by returning to an area where we are in agreement with our critics. Collins ends his paper by observing that, given the substantial progress in science studies over the last ten to fifteen years, it is time for the results to be 'cashed in' for their policy implications. We agree entirely, although it is important that policymakers are not 'short changed' in the process. We would, however, offer two further notes of caution. First, as with any 'applied' research, prior 'market research' is absolutely essential — in this case, to establish what are the needs of science policymakers. It cannot be assumed that what is most interesting from a sociological point of view is necessarily most relevant from a policy perspective. Secondly, sociologists and others in the science studies community will have to do rather more than weave fairy tales if they are to convince policymakers and scientists that their research has some validity and the results it yields some utility. In particular, their work must be capable of being defended in front of a scientific audience, where it will be judged in terms of criteria somewhat different from those normally employed in science studies.

Finally, on a more personal note, we warmly welcome the fact that Harry Collins, in discussing what constitutes 'the *best* way for science policy to make use of recent work in science studies'[57] seems at long last to be renouncing, implicitly at least, ultra-relativism. We applaud this courageous step 'forward', since science policy could potentially derive great benefit from the qualitative sociology of science if the two sides were to work more closely together than hitherto.

• NOTES

The authors wish to thank the Leverhulme Trust, which currently funds a three-year programme on research evaluation at the Science Policy Research Unit, for the financial support required in preparing this paper. They are also grateful to Andrew Barry, Diana Hicks, Ian Miles, Geoff Oldham, Keith Pavitt and John Ziman for helpful comments on an earlier draft of this paper.

1. No order of seniority implied (rotating first authorship).
2. J. Krige and D. Pestre, 'A Critique of Irvine and Martin's Methodology for Evaluating Big Science', *Social Studies of Science*, Vol. 15 (1985), 525–39, quote at 525.
3. R. Bud, 'The Case of the Disappearing Caveat: A Critique of Irvine and Martin's Methodology', *Social Studies of Science*, Vol. 15 (1985), 548–53, quote at 548.
4. H. M. Collins, 'The Possibilities of Science Policy', *Social Studies of Science*, Vol. 15 (1985), 554–58, quote at 554.
5. We are puzzled that, where we have focused on an entire field or discipline, this should still be classified by Collins as a unit with a 'non-cognitive' boundary.

6. Collins, op. cit. note 4, 554–55.

7. See J. Irvine and B. R. Martin, *Foresight in Science; Picking the Winners* (London: Frances Pinter, 1984), and Martin, Irvine and Turner, 'The Writing on the Wall for British Science', *New Scientist* (8 October 1984), 25–29.

8. Collins, op. cit. note 4, 555.

9. See J. Irvine and B. R. Martin, 'CERN: Past Performance and Future Prospects — II. The Scientific Performance of the CERN Accelerators', *Research Policy*, Vol. 13 (1984), 247–84.

10. Ibid., 281.

11. Collins, op. cit. note 4, 555.

12. Ibid., 554.

13. Ibid., 557.

14. Ibid.

15. See also *An Agenda for a Study of Government Science Policy*, a report prepared by the Task Force on Science Policy for the Committee on Science and Technology of the US House of Representatives, Ninety-Eighth Congress, Second Session (Washington, DC: US GPO, 1984). This lists some 140 science-policy related questions.

16. The points about 'non-convergence' raised by Collins, op. cit. note 4, 558, fn. 4, are discussed in the section below dealing with the criticisms of Moed and van Raan.

17. Bud, op. cit. note 3, 550 and 552.

18. For a response to Bud's criticism that we dismiss external (or non-scientific) benefits from high-energy physics as 'trivial', see the following section dealing with the comments of Krige and Pestre.

19. Bud, op. cit. note 3, 548.

20. See J. Irvine, B. R. Martin and M. Schwarz, with K. Pavitt and R. Rothwell, *Government Support for Industrial Research in Norway: A SPRU Report* (Oslo: Universitetsforlaget, Norwegian Official Publication NOU 1981: 30B, 1981).

21. Bud, op. cit. note 3, 548.

22. Ibid., 552.

23. Ibid., 552–53.

24. Krige and Pestre, op. cit. note 2, 526.

25. Ibid., 529.

26. J. F. Moed and A. F. J. Raan, 'Critical Remarks on Irvine and Martin's Methodology for Evaluating Scientific Performance', *Social Studies of Science*, Vol. 15 (1985), 539–47, quote at 544–45.

27. See Irvine and Martin, op. cit. note 9, 281.

28. H. Schmied, 'A Study of Economic Utility Resulting from CERN Contracts', *IEEE Transactions in Engineering Management*, Vol. EM-24 (1977), 125. This Study has recently been updated by Schmied and various colleagues at CERN — see M. Bianchi-Streit, N. Blackburne, R. Budde, H. Reitz, B. Sagnell, H. Schmied and B. Schorr, *Economic Utility Resulting from CERN Contracts (Second Study)* (Geneva: European Organization for Nuclear Research, CERN 84–14, 1984).

29. Irvine and Martin, op. cit. note 9, 281.

30. Bud., op. cit. note 3, 549.

31. See the discussion in J. Irvine and B. R. Martin, 'The Economic Effects of Big Science: The Case of Radio Astronomy', *Proceedings of the International Colloquium on Economic Effects of Space and other Advanced Technologies, Strasbourg, 28–30 April 1980* (Paris: European Space Agency, ESA SP-151, 1980).

32. B. R. Martin and J. Irvine, 'CERN: Past Performance and Future Prospects — I. CERN's Position in World High-Energy Physics', *Research Policy*, Vol. 13 (1984), 183–210; and Irvine and Martin, op. cit. note 9.

33. Even Schmied and his colleagues agree that 'The primary function of CERN is to carry out "very basic research" in particle physics, and the direct product of this research work is scientific knowledge or "culture"': Bianchi-Streit et al., op. cit. note 28, 1.

34. See, for example, M. Kaldor, 'Technical Change in the Defence Industry', in K. Pavitt (ed.), *Technical Innovation and British Economic Performance* (London: Macmillan, 1980), 100–25; K. Dickson, 'The Influence of Ministry of Defence Spending on Semiconductor Research and Development in the United Kingdom', *Research Policy*, Vol. 12 (1983), 113–20.

35. Krige and Pestre, op. cit. note 2, 530 (emphasis in original).

36. Ibid., 532.

37. For details, see B. R. Martin and J. Irvine, 'CERN: Past Performance and Future Prospects — III. CERN and the Future of World High-Energy Physics', *Research Policy*, Vol. 13 (1984), 311–42.

38. Krige and Pestre, op. cit. note 2, 536 (emphasis added).

39. Moed and van Raan, op. cit. note 26, 542.

40. See the distinction between 'quality', 'importance' and 'impact' made in B. R. Martin and J. Irvine, 'Assessing Basic Research: Some Partial Indicators of Scientific Progress in Radio Astronomy', *Research Policy*, Vol. 12 (1983), 61–90.

41. Ibid., 65–74.

42. Moed and van Raan, op. cit. note 26, 544, and again at 541.

43. Taken from Tables 3 and 4 in J. Irvine and B. R. Martin, 'Assessing Basic Research: The Case of the Isaac Newton Telescope', *Social Studies of Science*, Vol. 13 (1983), 49–86.

44. Martin and Irvine, op. cit. note 37, 87, fn 55.

45. This may explain the claim by Moed and van Raan (op. cit. note 26, 546) that this citations-per-paper indicator does not allow for differences in scale of research activity. While this claim may be true for fields characterized by localized, non-interacting communities with differing propensities to engage in self-citation and in-house citation, for a reasonably internationally homogeneous field like high-energy physics, the citations-per-paper figure for a particular facility is much less dependent on its publication rate.

46. Moed and van Raan are wrong in concluding that we have assumed that Eastern-bloc papers refer *exclusively* to other Eastern-bloc papers (ibid., 545). What we have allowed for is the fact that Eastern papers lose *far more* citations from unscanned journals that do Western ones, and we did this by taking a sample of references in Eastern papers and examining the proportions given to Eastern and Western papers.

47. See Martin, Irvine and Turner, op. cit. note 7.

48. This was a major point of discussion in the UK Advisory Board for the Research Council's Science Policy Study 1983/84, part of which is described in B. R. Martin, J. Irvine and D. Crouch, *Science Indicators for Research Policy: A Bibliometric Analysis of Ocean Currents and Protein Crystallography* (Brighton: University of Sussex, Science Policy Research Unit Occasional Paper No. 23, 1985).

49. For details, see ibid.

50. Moed and van Raan, op. cit. note 26, 545.

51. Ibid.

52. See H. F. Moed, W. J. M. Burger, J. G. Frankfort and A. F. J. van Raan, *On the Assessment of Research Performance: The Use of Bibliometric Indicators* (University of Leiden: Research Policy Unit, 1983).

53. Moed and van Raan, op. cit. note 26, 546–47.

54. Ibid., 544.

55. CERN member states like Britain expect the bulk of their high-energy physicists to use CERN facilities (only very limited resources are available to fund experiments at US laboratories, for example).

56. See J. Irvine and B. R. Martin, 'What Direction for Basic Scientific Research?', in M. Gibbons, P. Gummett and B. M. Udgaonkar (eds), *Science and Technology Policy in the 1980s and Beyond* (Harlow, Essex: Longman, 1984), 67–98.

57. Collins, op. cit. note 4, 554 (emphasis added).

Authors' address: Science Policy Research Unit, University of Sussex, Mantell Building, Falmer, Brighton, East Sussex BN1 9RF, UK.