Assessing basic research

Some partial indicators of scientific progress in radio astronomy

Ben R. MARTIN and John IRVINE *

Science Policy Research Unit, University of Sussex, Brighton BN1 9RF, UK

Accepted for publication September 1980

As the costs of certain types of scientific research have escalated, and as growth rates in overall national science budgets have declined, so the need for an explicit science policy has grown more urgent. In order to establish priorities between research groups competing for scarce funds, one of the most important pieces of information needed by science policy-makers is an assessment of those groups' recent scientific performance. This paper suggests a method for evaluating that performance.

After reviewing the literature on scientific assessment, we argue that, while there are no simple measures of the contributions to scientific knowledge made by scientists, there are a number of 'partial indicators' - that is, variables determined partly by the magnitude of the particular contributions, and partly by 'other factors'. If the partial indicators are to yield reliable results, then the influence of these 'other factors' must be minimised. This is the aim of the method of 'converging partial indicators' proposed in this paper. We argue that the method overcomes many of the problems encountered in previous work on scientific assessment by incorporating the following elements: (1) the indicators are applied to research groups rather than individual scientists; (2) the indicators based on citations are seen as reflecting the impact, rather than the quality or importance, of the research work; (3) a range of indicators are employed, each of which focusses on different aspects of a group's performance; (4) the indicators are applied to matched groups, comparing 'like' with 'like' as far as possible; (5) because of the imperfect or partial nature of the indicators, only in those cases where they yield convergent results can it be assumed that the influence of the 'other factors' has been kept relatively small (i.e. the matching of the groups has been largely successful), and that the indicators therefore provide a reasonably reliable estimate of the contribution to scientific progress made by different research groups.

In an empirical study of four radio astronomy observatories, the method of converging partial indicators is tested, and several of the indicators (publications per researcher, citations per paper, numbers of highly cited papers, and peer evaluation) are found to give fairly consistent results. The results are of relevance to two questions: (a) can basic research be assessed? (b) more specifically, can significant differences in the research performance of radio astronomy centres be identified? We would maintain that the evidence presented in this paper is sufficient to justify a positive answer to both these questions, and hence to show that the method of converging partial indicators can yield information useful to science policy-makers.

* No order of seniority implied (rotating first authorship). The authors are currently Fellows of the Science Policy Research Unit where they work on a range of issues connected with policies for basic research. They wish to acknowledge the support of the Social Science Research Council in carrying out this research. They also want to thank their colleagues at the Science Policy Research Unit, particularly Geoff Oldham, Keith Pavitt, and Roy Turner, for their help and comments. The conclusions are, however, the responsibility of the authors alone, and are not necessarily shared by their colleagues. Norman Dombey wishes to dissociate himself completely from the results of this publication.

Research Policy 12 (1983) 61–90
North-Holland Publishing Company

1. Introduction – The need for an assessment of basic research

Few would now dispute the necessity for some assessment of the output from basic research. As the scale of basic scientific research has grown, so the need for an explicit and systematic science policy has increased – a policy that can only be sensibly constructed on the basis of an evaluation of the output from that research. When basic research was little more than a hobby for gentlefolk, and even as recently as 50 years ago when the annual equipment budget for an entire university department like the Cavendish Laboratory at Cambridge University amounted to only a few hundred pounds, it was entirely acceptable to leave questions concerning the organization and funding of science to scientists. There was no need for a policy establishing priorities across the sciences, nor for an evaluation of the performance of scientists; any regulation of scientific affairs could be left to the scientists themselves. Now, however, the resources devoted to basic research are so substantial that its practitioners can no longer hope to escape the scrutiny of those seeking to know whether public funds are being spent in areas that yield a good rate of return (whatever form those returns may take). In Britain, the budget of the Science Research Council (SRC) for the financial year 1978–1979 amounted to well over £150 million. Although some of this was allocated to engineering research, and some to certain, more applied areas of biology, chemistry, and physics, the great bulk went to basic research – that is, research carried out with the primary purpose of increasing scientific knowledge rather than creating technological, social, or economic benefits.

The need for a policy for basic research, and hence for analytical tools to assess such research, has been greatly sharpened by the problems of economic recession in the 1970s. Industrialised nations have been forced to look more carefully at all areas of state expenditure. Whereas in the 1950s and 1960s economic growth rates generally permitted public expenditure in one area such as scientific research to grow, and even to grow quite rapidly, without affecting other areas, now an increase in one area of public expenditure tends only to be possible at the expense of cut-backs in others. Those scientists who work in high-cost specialties such as high-energy physics, space research, and astronomy – where new facilities can cost tens or even hundreds of millions of pounds – and who argue for continued real increases in their level of funding, must now be willing to accept wider public examination of their research work in order that one may ascertain what benefits are associated with it, and how these benefits compare with those from other areas competing for resources. It is no longer sufficient to assert that a particular project promises certain benefits or "good science"; instead, it must be demonstrated that the project is likely to yield *greater* benefits than any of the competing alternatives. Such a judgement can only be arrived at on the basis of an assessment of the benefits associated with all the various areas of basic research, and the establishment of a set of relative priorities between them in one comprehensive science policy.

For policy-makers, there are four main sets of decisions that need to be made in the allocation of resources to basic research.

(1) How much overall should be spent on basic research compared with other areas of public expenditure?
(2) How should this overall research budget be distributed between the different disciplines, each with competing claims for funding?
(3) How much should be allocated to the different types of scientific institutions? In Britain, for example, what proportions should be spent on international facilities, on Research Council establishments, and on university research?
(4) How much should be allocated to each research centre, group, or individual, working within a discipline?

Some consideration was previously given to these questions in the debate on criteria for scientific choice during the 1960s (cf. Shils [49]). In particular, Weinberg [53, p. 163] proposed certain criteria on which decisions as to the distribution of resources to scientific research should be based, distinguishing between "internal" and "external" criteria:

> Internal criteria are generated within the scientific field itself and answer the question: How well is the science done? External criteria are generated outside the scientific field and answer the question: Why pursue this particular science?

In the debate stimulated by Weinberg's proposals, participants tended to concentrate more on the policy issues associated with applying *external* criteria to decisions about scientific funding, and hence to focus on the first two of the four questions identified above. Much attention was given to such factors as the contributions of science in general, or of a particular scientific discipline, to education, to technology, and to economic growth; and governments have in recent years begun to take account of some of these external factors in arriving at decisions on funding. However, internal criteria cannot be ignored, particularly in the case of more basic scientific research. It is with these that we are primarily concerned here.

In this paper, we present a framework for assessing the relative contributions to scientific knowledge made by different research groups in the same discipline. This has been developed as part of a study of a small number of major, basic-research centres. [1] The main reason for focussing on the centre as the unit of analysis rather than the individual scientist is that over half the annual SRC expenditure goes to support research centres rather than individual scientists or projects. [2] The principal objective of this work is to provide information which may facilitate the taking of decisions of the third and, in particular, the fourth type identified above. In deciding the distribution of resources between research centres, one of the most relevant pieces of information is the past performance of those centres. "Track record" is by no means the *only* factor determining future research performance (there are others like the "ripeness" or potential for exploitation of new work), but it is undeniably one of the most important. If other factors are equal, or alternatively if they are completely indeterminate, then one has to assume that, on average, a new scientific project is *more* likely to be carried out well by a group with a record of successful research over the recent past than by one with a less distinguished record. We are concerned in this paper with the question of the extent to which past performance can be reliably and satisfactorily evaluated. The questions to which we have addressed ourselves take the following form: to what extent can the outputs of research groups or centres working in the same discipline be compared? Is it possible to assess whether one research group has contributed more to scientific knowledge than another, and, if so, how?

The structure of the paper is as follows: the first sections discuss the nature of basic research, and the outputs from it; this is followed by a critical review of the literature on previous uses of various measures of science, drawing out the severe methodological and conceptual problems associated with each measure; on the basis of this, it is shown that, once some consideration is given to the question of what is actually being measured, *careful* use of a number of "partial indicators" of scientific progress can yield information on the relative performance of groups of scientists; the later sections then detail our empirical work in which several partial indicators are combined to assess the performance of a number of large, basic-research centres working in the field of radio atronomy.

2. The nature of basic research and its outputs

While assessing the output from basic research may be extremely difficult, there can be no doubt that there is an output of some kind. This may take the form of new scientific knowledge (theories, empirical findings, and so on), new scientific problems, or new practical ideas or problems (cf. Freeman [14]). In short, there is a flow through basic science of information generally embodied in research publications or in people. Borrowing a model from the conceptual vocabulary of economists, we can usefully conceive of science as an "input–output" process (cf. e.g. Moravcsik [38]). Such a model of science is depicted in fig. 1. While there are few major difficulties here in identifying suitable input measures, there are severe conceptual as well as methodological problems associated with finding appropriate output measures. These stem from the intangible nature of much of the output or "product" of basic research activities. Indeed, the very nature of the "product" depends on our philosophy of science (for instance, on whether we assume scientific knowledge is cumulative in nature, or whether its advance is better represented as a series of revolutionary transformations – cf. Kuhn, [25]), as well as on our approach to the sociology of scientific institutions. Part of the problem is that there are many ways in which contributions to scientific progress are made. While a few scientists are responsible for making major "discoveries", most make relatively small incremental additions to our knowledge (perhaps in the form of more precise measurements). Both types of contribution are obviously essential to scientific progress. In addition, scientists who are primarily teachers, administrators, or technicians, all play crucial roles in scien-

[1] The project aimed to assess the performance of five major centres supported by the SRC. These were the Daresbury and Rutherford high-energy physics laboratories, the radio astronomy observatories at Cambridge and Jodrell Bank, and the optical astronomy facilities at the Royal Greenwich Observatory.

[2] The CERN, Rutherford, Daresbury, Appleton, and ILL Grenoble Laboratories, together with the two Royal Observatories, accounted for £87 million of the £157 million spent by the SRC in the financial year 1978–79, while several other major research groups, including the radio astronomy observatories at Cambridge and Manchester Universities, were also supported mainly through consolidated block grants rather than through grants for specific research projects.

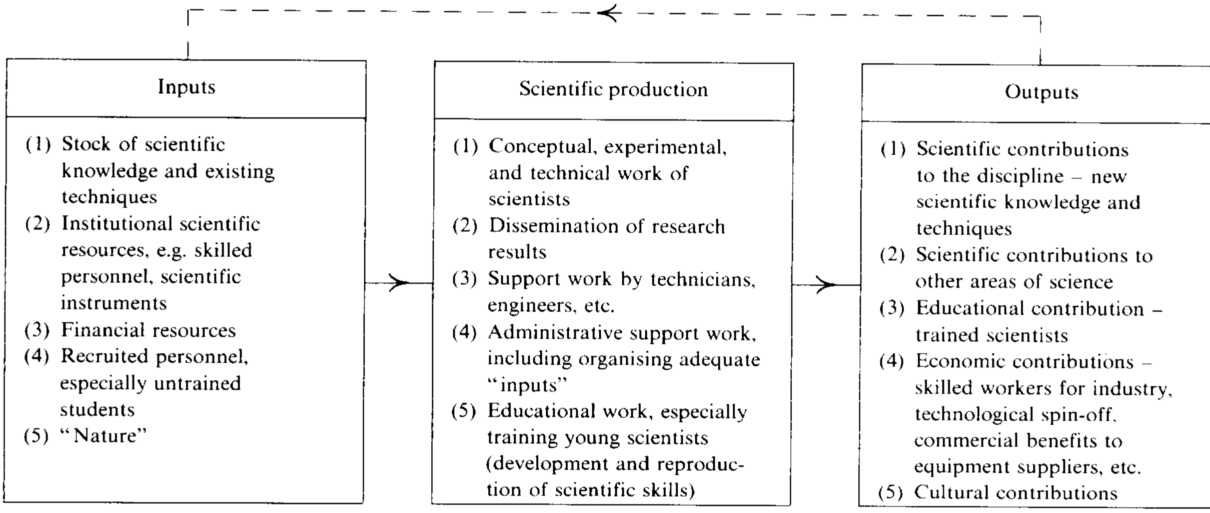

Fig. 1. An input–output model of a basic research discipline.

tific development (cf. Lawani [28, p. 26]). Although there are other such facets to scientific activity, we shall be concentrating here on assessing contributions to scientific knowledge, [3] since these are most directly related to the primary goal of basic research.

There are a number of possible yardsticks for assessing the contributions to scientific knowledge made by individuals or groups of scientists. These include the number of scientific publications produced in a given period or for a given volume of resources, the number of times these publications are cited in other articles or books, the evaluation by scientific peers of the importance of the published work, the number of "discoveries" or other major advances in knowledge, and the recognition afforded to the authors of the publications (in the form of honours or prizes, for example). Some of these measures can be used fairly readily on a small scale, but it is difficult in practice (and perhaps invalid theoretically) to extend such indices of performance across a wide range of disciplines and countries, or over an extended timescale (cf. Freeman [14, p. 4]). It is the contention of this paper that, although no *absolute* quantification of basic research is possible, one can make valid and useful *comparisons* between the scientific performance of different research groups, provided that careful thought is given both to the choice of groups for comparison, and to the question of what the various indicators of scientific performance are actually measuring.

[3] We examine the external benefits from radio astronomy in another paper (Irvine and Martin [22]). This considers both technological spin-off and the migration from radio astronomy of trained students, who take with them skills that subsequently prove useful in a variety of high-technology jobs. However, the main focus of our work has been on contributions to scientific knowledge, because nearly all the members of the research groups whom we interviewed gave this as the primary justification for continued financial support of their research, rather than other outputs such as contributions to education or to technology.

3. Scientific activity, scientific production, and scientific progress

Before looking at the various measures of output from basic research, it is helpful to separate out different aspects of the performance of scientists, and to make a distinction between scientific activity, scientific production, and scientific progress (cf. Moravcsik [37, p. 268]). The first, scientific activity, is concerned with the consumption of input resources, and is therefore related to such factors as the number of scientists involved, the expenditure on their research, the percentage of their time spent on research, and the number of support staff (for example, in the form of technicians or administrators). The second, research production, refers to the extent to which this consumption of resources creates a body of scientific *results*. These results are usually manifest in the form of research communications, although scientists do communicate through other informal channels such as letters, seminars, and personal conversations. [4] The third, scientific progress, refers to the extent to which scientific activity actually results in substantive contributions to scientific *knowledge* (as judged by other scientists).

The problems of assessing scientific progress stem partly from the fact that these contributions to knowledge are not always regarded as cumulative, which means that such concepts as the "quantity" and "quality" of research (terms apparently implying absolute characteristics) may be misleading. As we shall see below, while some of the output indicators are fairly clearly linked to scientific production, their links with scientific progress are more complex and problematic. Yet it is with indicators of scientific progress that we must be most concerned if we are to evaluate the extent to which scientists succeed in fulfilling the primary goal of basic research, the production of new scientific knowledge.

4. Publications

The first output measure we shall consider is based on scientific publications. Several authors have made use of the distinction between the "quantity" and "quality" of research contributions (e.g. Lawani [28, p. 26]). The former has generally been measured in terms of numbers of publications, and has been the basis for several studies of the growth of science, [5] and of the performance of groups of scientists (e.g. Chang and Dieks [4]). [6] However, such studies have often exhibited a lack of conceptual clarity as to what the number of publications actually measures. While it may be regarded as a reasonable measure of scientific production, [7] its status as an indicator of scientific *progress* is uncertain. Some authors claim to have found a high correlation between numbers of publications and the overall "quality" or "merit" of the papers (e.g. Clark [6]), or the "eminence" of their authors (e.g. Price [47], p. 41]). [8] However, as we shall see below, the relationship of these variables to scientific progress is not always straightforward. Indeed, for some scientists, or groups of scientists, the correlation between "quantity" and "quality" is small or even zero (cf. Smith and Fiedler [50, p. 228]).

The problem is that each publication does not represent an equal contribution to science. Some "mass-producers" of publications make very little scientific progress, while other "perfectionists" achieve in a few publications very significant scientific progress (cf. Cole and Cole [7, p. 382]). Various attempts have been made to circumvent this problem, for example, by "weighting" some publications differently from others, [9] but this has tended to be done without any adequate theoretical basis for the choice of weights (cf. Bayer and Folger [2, p. 382]). Publication counts by themselves fail to "distinguish between the fluency of genius and the loud noises of empty vessels" (*Nature* [43]); they give too much emphasis to the "operator" who produces quantity rather than quality (cf. Bayer and Folger [2, p. 382]. But how can one measure "quality"?

The task of assessing the relative quality of publications has not been made easier by the explosive growth in scientific literature. Over 15 years ago, Maddox [30, p. 15] remarked that, "in so generally flaccid a literature it is difficult to tell

[4] These informal channels of communication may be very important, particularly in the initial stages of disseminating new scientific knowledge. However it has been assumed here (although this does need empirical investigation) that much of the information passing through these channels ends up in the final depository of scientific publications, and hence that only research publications need be assessed (cf. Moravcsik [38]).

[5] See Gilbert and Woolgar [20, pp. 279–83] for a summary of some examples.

[6] Numbers of publications are also apparently used by some administrators responsible for allocating resources in science. A recent article in *Nature* [45] refers to a professor of chemistry (and Nobel Laureate) being threatened with expulsion by his university for "inadequate productivity"!

[7] But see note 4.

[8] Price argues that "on the whole there is, whether we like it, or not, a reasonably good correlation between the eminence of a scientist and his productivity of papers. It takes persistence and perseverance to be a good scientist and these are frequently reflected in a sustained production of scholarly writing." Similarly, Cole and Cole [7, pp. 387–8] claim that "high producers *tend* to publish the more consequential research...(because) engaging in a lot of research is in one sense a 'necessary' condition for the production of high quality work...(and because) the reward system operates in such a way as to encourage the creative scientists to be productive and to divert the energies of the less creative scientists into other channels".

[9] See Bayer and Folger [2, p. 382] for a summary of such attempts.

which papers are substantial contributions to understanding, and which are but trivial documents", and the situation has certainly not improved since that time. Some discussion has been given to the possibility of weighting publications according to the journal in which they appear (e.g. Garfield [18]). However, there is evidence that even the prestigious journals are not always very discriminating (cf., for example, Lawani [28, p. 26]). Hence, attempts to attach a "quality index" or "impact factor" (Garfield [17, p. 474]) to journals fail to confront the problem of the wide variation in quality *within* each journal (cf. Smith and Fiedler [50, pp. 230–31]). The two principal methods which have been used to obtain an indicator of the quality of publications involve peer evaluation and citation analysis. [10] How these are linked with scientific progress is explored in subsequent sections.

Without some allowance being made for variations in quality, a simple count of publications, while it may provide a measure of scientific production, is not a *measure* of scientific progress. It is what we shall term a *partial indicator* [11] *of scientific progress* – that is, a variable determined *partly* by (a) the level of scientific progress made by the individual or group, but also influenced by (b) other factors, such as various social and political pressures. These include the publication practices of the employing institution (current *and* previous institutions – see Crane [11, p. 703]), the country, and the research area. Another factor is the emphasis placed on numbers of publications for obtaining promotion, tenure, or grants. [12] The situation is complicated by the fact that these 'other factors' not only vary between scientists or groups of scientists, but may also change with time. [13] Moreover, the question of whether (a) dominates over (b) is problematical. It cannot be *assumed,* a priori, that (b), compared to (a), is relatively insignificant, or even that it comprises a set of entirely random influences whose effects can be taken to cancel out either for all large aggregations of scientists or for an extended time-period. (Some of the effects may be random, but others will be systematic in nature, depending on the particular set of social and political circumstances.) The relative importance of (a) and (b) on publication counts can only be established empirically. In short, we have to consider *why* scientists publish papers, and to realise that they do so not only to present valuable results, but also for social, political, and career reasons.

Treating publication counts as a partial indicator, rather than a measure, of scientific progress helps us to see why it may be dangerous to rely on numbers of publications for assessing the scientific output of an individual, or even of a small group, since it is in these cases that there may be wide, and possibly unpredictable, variations in the relative effects of (b). With larger groups, however, it may be possible to select carefully matched groups for comparison so that we minimize some of the effects of (b) – for example, by selecting groups working in the same specialty, under similar systems of organization and funding, and publishing similar types of papers in the same journals. We *may* then find that the effects of (b) relative to (a) can be reduced sufficiently to make publication counts a useful indicator of scientific progress. Some preliminary empirical results on this are presented in the latter part of this paper.

[10] Citation analysis has already found a variety of uses, particularly in the United States, where the National Science Foundation uses it to assess, for example, its funding of chemistry departments; where various universities employ it to decide cases of promotion and tenure; and where citation evidence was produced in court to prove that a woman denied tenure was as good as, or better than, two men promoted over her (Wade [52, p. 429]).

[11] To a certain extent the term "partial indicator" is tautological in that use of the word "indicator" already implies a partial or incomplete measure. However, in view of the tendency of some users of indicators to forget this partiality or incompleteness in the nature of indicators, and the consequent danger of confusing the conceptual with the ontological, i.e. equating the mental construct with "reality", it is worth emphasizing the partiality. Consequently, the term "partial indicator" has generally been used throughout the paper, although where repetition of the word "partial" would be labouring the point unnecessarily, this has been abbreviated to "indicator".

[12] As the use of publications to account for time and money spent on research has grown, so the pressures to publish have increased correspondingly, resulting in a state of "literature inflation" – a vicious circle in which the more papers that are written, the less they count for, and the greater is the pressure to publish more (cf. Margolis [31, p. 1218]).

[13] Cf. e.g. Sullivan et al., [51, pp. 182–84], who find that the productivity of certain theoretical physicists has quadrupled over a 20-year period, while that of experimentalists working in the same specialty has stayed constant.

5. Citations

The aim of citation analysis is generally held to be that of injecting a "quality factor" into the evaluation of scientific publications. A fundamental assumption of most previous use of citation analysis is that the impact of a paper lies in its influence on subsequent research papers, and that each instance of such influence will manifest itself by the influenced paper referring to the influencing paper (cf. Moravcsik [38]). On the basis of this somewhat challengeable assumption, the more enthusiastic devotees of citation analysis assert that the number of citations earned by a paper represents the "quality" of that paper (cf. e.g. Cole and Cole [7, p. 379]). [14] Various pieces of evidence have been adduced to support this claim; for example, Clark [6] finds a high correlation between citations and quality, and concludes that a citation count is the best available indicator of the "worth" of research work. Another kind of evidence is said to come from Nobel Laureates, the majority of whom are amongst the top 0.1% most cited authors (cf. Garfield [19, p. 485]), giving rise to the suggestion of "predicting Nobel Prize winners" by citations (Garfield [16, p. 671]). On the basis of this and other evidence, [15] it has been concluded by some that citations are "the best practical indicator of the worth of research" (Porter [46, p. 257]). How justified is this claim?

As most citation analysts are ready to admit, counts of citations face a number of problems, some technical, some more substantive (cf. Cole and Cole [9, p. 24]). Technical problems stem from the fact that the *Science Citation Index,* which is generally used in these studies, does not provide a totally accurate reflection of citation structure (cf. Moravcsik [38]). For example, citations to multi-authored publications are listed only under the first-named author. Since there appears to be a high correlation between citation counts obtained in this way and *total* citation counts (obtained by looking up a list of all the papers by a researcher regardless of whether they were the first-named author), it is generally assumed that using only first-named authors makes little difference (cf. e.g. Cole and Cole [9, pp. 32–34]. While this may be true on average for *most* scientists, it can make a crucial difference to the citation counts of *some* individuals (cf. Lindsey [29]), particularly penalizing those whose work involves a great deal of collaboration. Hence, in comparisons between groups of scientists who engage in differing degrees of collaboration in their research, use of first-named author citations may lead to significant systematic errors, and in the empirical work described below total citations are used for this reason. A second technical problem is that of individuals who are listed under more than one name in the *Science Citation Index,* perhaps because they sometimes use one initial and sometimes two, or alternatively because of a change of name on marrying; these, however, can generally be checked where such variations are suspected. Similarly, some names prove "difficult" for citing authors to spell, and again one may have to check under several different possibilities in the *Index.* Yet another problem is that of two authors with the same name and initials, but this can generally be dealt with provided the research interests of the scientists and the journals in which they customarily publish are known. In the empirical study described later, one difficulty concerned preprints which are listed together in the *Science Citation Index* under "in press". In a small number of cases of collaborative papers it was not possible to establish conclusively whether the cited preprint was the product of the research group under examination, or that of another research centre, but in our study such instances were very rare (less than 0.5% of all citations). Finally, some citing authors make clerical errors in recording page numbers, volume numbers, and journal names. These can generally be identified through comparison with a complete and accurate list of publications for the individual concerned. In short, most of the technical errors can be overcome with care and diligence; and this effort would seem well warranted if Porter's [46, p. 264] estimate that, if uncorrected, these errors could be as large as 25%, is anywhere near accurate.

The substantive problems associated with citation analysis are more complex and severe. First, there is the case of high-quality work that is initially resisted or neglected, perhaps because it is "ahead of its time", and which therefore accu-

[14] In a later paper, Cole and Cole [9, pp. 23–24]) make a distinction between "absolute quality" (i.e. papers embodying "absolute truth") and "socially determined quality", but they still maintain that citations "are an adequate measure of the quality of work socially defined".

[15] See e.g. Bayer and Folger [2]; Cole and Cole [7]; Myers [41].

mulates few citations. The assertion that such occurrences are rare (cf. e.g. Cole and Cole [9, pp. 24–25]) needs more empirical justification than the quoting of a few famous instances such as Mendel, accompanied by a claim that these occur most infrequently. Secondly, poor quality papers may be frequently cited, because they are either controversial or "mistaken" (cf. e.g. Janke [23, p. 892]). The response of Cole and Cole [10, pp. 32–33] to this criticism is that,

> Much important work turns out in historial retrospect to have been "incorrect"... A paper which is important enough to receive a large number of citations is probably a significant contribution. [16]

While there is more than a little validity to this argument, it illustrates the conceptual confusion surrounding the use of the term "quality", the characteristic which citations are often claimed to measure.

Thirdly, there is a problem in dealing with the small number of extremely important papers (such as those by Einstein) that present new fundamental ideas. Once these ideas have been accepted and fully integrated into the body of scientific knowledge, the original papers may be seldom cited. Cole and Cole [9, p. 31] admit that this may lead to errors for individual papers, but assert, again with little apparent supporting evidence, that this is not a problem when dealing with a number of papers or with scientists. Fourthly, a high-quality paper in a small, narrow, or highly specialized field may produce few citations – fewer than a similar quality paper in a more popular field (cf. McGervey [35, p. 30]). On the basis of comparisons between solid-state physics and the smaller specialty of high-energy physics (although this would appear to be a far from typical "small" field), Cole and Cole [9, pp. 28–29] conclude there is no evidence for such an effect, though they do admit that it might be a problem for very small fields. [17] Other authors dispute this claim by the Coles, arguing that, if citations are used as a measure of quality, one must take account of the size of population of articles that might potentially refer to the work in question (cf. e.g. Sullivan et al. [51, pp. 195–96]) – in other words, the numbers of citations must be "normalized" if one is attempting to compare research in different specialties, or to compare research in the same specialty at different times.

A fifth factor influencing the relationship between numbers of citations and quality is the effect of self-citation, the incidence of which may be partly related to the size or the state of development of the specialty, [18] although it is also structured by a number of other factors including the degree of "openness" of the individual or group concerned to work carried out elsewhere. Matheson's [33, p. 208] study of British university chemistry departments reveals that self-citation rates can vary considerably (by a factor of two) between different research groups working in the same field, and this suggests that, at least for groups practising a high degree of self-citation, the number of citations received may be at least partially determined by the number of publications they produce. In consequence, numbers of citations may reflect "quantity" as well as "quality" of publications.

Another problem with the use of citations as a measure of quality is that certain *kinds* of papers are more frequently cited than others of similar quality. For example, authors in theoretical physics

[16] The Coles [10, pp. 32–33] continue: "Why should a large number of scientists waste their time pointing out a trivial error? In fact they do not. Papers which are trivial and receive critical citations will not accumulate large numbers of citations."

[17] Citation patterns within a specialty *do* change considerably with the size of the specialty. For example, Meadows and O'Connor [34] show that the average number of references per paper jumped from 7.1 to 9.9 during the first year of research in the new and rapidly growing field of pulsar research. Similarly, Sullivan et al. [51] note a doubling in the average number of references per paper over a 20-year period in the research area of weak interactions, a sub-field of high-energy physics. In contrast, it is argued by some (e.g. Lawani [28, p. 30]) that the size of the research field will not greatly influence the relationship between the quality of a publication and the number of citations it receives because, although there are fewer papers to cite from in a small field, each paper is also more prominent than one of similar quality in a broader field. However, there must be doubts as to whether these two opposite size-effects do exactly cancel out, and hence whether the relationship between quality and citation does "scale up" exactly in this way.

[18] For example, Meadows and O'Connor [34, p. 97] find that the percentage of self-citations dropped from 15% of the total citations to 10% over the first year of pulsar research as more, potentially citable work by other authors became available.

frequently refer to accounts in the mathematical literature of the development of a new technique simply as a means of shortening their own exposition. Similarly, in chemistry and biology, papers describing a standard technique [19] may be highly cited regardless of whether or not the first published description of the technique constitutes a high-quality piece of research (cf. *Nature* [44, p. 699]). Hence, the number of citations a paper receives may reflect its nature as well as its quality.

Finally, the use of citation counts is complicated by the "halo effect" – the fact that, when there is a choice of publications to cite, the author of a paper is more likely to refer to the work of eminent scientists (cf. Smith and Fiedler [50, p. 229]). Scientists who have gained a reputation by publishing significant research in the past then enjoy a halo effect in that all of their research, even if some of it is of lower quality, gains additional attention. Cole and Cole [9] admit that, if one considers the total number of a scientist's citations, some will be due to the halo effect, but they argue that the size of the effect should be directly related to the significance of the scientist's research and that this does not therefore affect the use of citations as a measure of quality. There are, however, at least two major weaknesses in this argument. First, the quality of currently cited work may not be as good as that of the previous research that determines the size of the halo effect (cf. Croom [12, p. 1173]). Secondly, the relationship between the halo effect and the quality of past research may not be a *linear* one, with the result that analysis based on numbers of citations could *exaggerate* the differences between high quality and low quality research. [20] For example, the halo effect enjoyed by a Nobel Prize winner is probably far larger than that accorded to the winner of a British Royal Society medal (often for work of not greatly inferior quality) because of the tremendous publicity given by the semi-scientific and popular press to the former. [21]

Underlying all these problems with the use of citations as a measure of quality is our ignorance of the reasons *why* authors cite particular pieces of work and not others. The problems described above arise principally because simple citation analysis presupposes a highly rational model of reference-giving, in which citations are held to reflect primarily scientific appreciation of previous work of high quality or importance, and potential citers all have the same chance to cite particular papers (regardless of who the authors are, where they published, what language they wrote in, etc. – in other words a "free market" in scientific ideas). The model also assumes that the norms of citation behaviour are largely unaffected by any "external" pressures (including the awareness of scientists that citation analyses may be used to assess their scientific performance), and that all citations are of equal value or intent. Such a model is obviously a grossly oversimplified and possibly highly misleading representation of reference-giving; the activity of reference-giving is, like other aspects of scientific work, a social activity, and an author's citation practices can only be fully understood if they are related to individual cognitive interests, institutional position, and social and political goals. Before we look to see how this can be done in practice, certain conceptual distinctions must be made in order to ascertain what the numbers of citations may or may not indicate.

6. The quality, importance, and impact of publications

Previous use of citation measures has, as we have seen, mostly exhibited a lack of conceptual clarity as to what the citation rates measure. To overcome these problems, some authors have attempted to distinguish between the "quality" and the "impact" of publications, arguing that, even if citations do not provide an indicator of the quality of a paper, they do at least reflect its impact on the scientific community (e.g. Garfield [15, p. 30]).

[19] O.H. Lowry, the most highly cited scientist, regularly receives several thousand citations each year (several times more than even the most eminent of scientists), the bulk of which refer to a technique for protein determination. Given that this technique is now so standard, one wonders whether all those who cite this work have in fact read the relevant paper before they make use of it.

[20] This may partly account for the controversial conclusion of Cole and Cole [8, p. 368] that "only a few scientists contribute to scientific progress".

[21] Cf. Inhaber and Przednowek [21, p. 33]: "In practice, the semi-scientific and popular press considerably distort the visibility of the different forms of recognition. This may produce a gap between the perceived and true quality of scientific work."

However, few have specified exactly to what each of these concepts refers. We would argue that it is necessary to distinguish between, not two, but three concepts – the "quality", "importance", and "impact" of the research described in a paper – if we are to understand what, if anything, citation-rates measure (cf. Lawani [28, p. 27]). The first of these concepts refers to the research itself, while the latter two are more external, referring to the relations between the research and other research areas, and describing the strength of the links to, or the implications for, other research activities.

"Quality" is a property of the publication and the research described in it. It describes how well the research has been done, whether it is free from obvious "error", how aesthetically pleasing the mathematical formulations are, how original the conclusions are, and so on. But quality is still relative rather than absolute, and it is socially as well as cognitively determined; it is not just intrinsic to the research, but is something judged by others who, with differing research interests and social and political goals (i.e. different cognitive and social "locations" – cf. Martin [32]), may not place the same estimates on the quality of a given paper. Even the same individual may evaluate the quality of a paper differently at different times because of progress in scientific knowledge and shifts in his or her location.

The "importance" of a publication refers to its *potential* influence on surrounding research activities – that is, the influence on the advance of scientific knowledge it would have if there were perfect communication in science (in short, if there were the completely "free market" of scientific ideas mentioned above). However, there are "imperfections" in the scientific communications system, the result of which is that the *importance* of a paper may not be identical with its *impact.*

The "impact" of a publication describes its *actual* influence on surrounding research activities at a given time. While this will depend partly on its importance, it may also be affected by such factors as the location of the author, and the prestige, language, and availability, of the publishing journal (cf. Dieks and Chang [13, p. 247]).

Having distinguished these three concepts, we must next examine how they are related to each other, and to the notion of scientific progress described earlier. The importance and the impact of a publication are determined by the relations between the research described in it and the surrounding specialty, and hence by the quality of the former and by certain characteristics of the latter – for example, the level of research activity it exhibits. This level of activity in turn reflects scientists' perceptions as to which specialties deal with the most "fundamental" questions (Cole and Cole [10, p. 32]), which find it easiest to attract research funding, which are more likely to attract high quality students, and so on – in other words, factors dealing with the relationship between the narrow research area and the wider research domain. It seems plausible to suggest that a high quality paper in an active specialty will generally have a greater importance than a similar quality paper in a dormant or declining one. [22] If so, we can see that the "quality" of a publication need not be synonymous with its contribution to scientific knowledge; for example, a high-quality paper in a stagnating field may contribute little to the advance of scientific knowledge in general. Nor is the "importance" of a paper necessarily an indication of the size of its contribution; important papers can go unnoticed if authors express themselves poorly, if they publish in a journal with a restricted circulation, [23] or if they have not previously been particularly prominent in the scientific community. [24] It is the "impact" of a publication that is most closely linked to the notion of scientific progress – a paper creating a great impact represents a major contribution to knowledge *at that time* (although its impact may of course alter with time).

Is it possible to obtain any absolute or direct measure of the quality, importance, or impact of a publication? The short answer is "No". These factors are not absolute (in the sense of holding the same value for all people and all time) but relative, varying over time and according to the

[22] Unless the work is so exceptional that it revitalises a previously stagnating research area, sending reverberations far beyond its boundaries to neighbouring areas.

[23] Cf. Lawani [28, p. 31], who notes that entomological articles published in West Africa are cited less than those published overseas.

[24] Cf. Whitley [55, p. 230], who finds some evidence for this "bandwagon" or "halo" effect: "once a paper by a man is cited there may be some incentive to cite other papers by him. Partly it may be due to the intrinsic quality of the papers and partly to the high visibility of the man and his work."

cognitive and social location of the assessor. Moreover, they cannot be evaluated directly, but only apprehended indirectly through the perceptions of other scientists (by peer evaluation), or partially inferred from the social practices (for example, citation practices) of scientists. The number of citations to a publication is not a direct reflection (a "measure") of its quality or importance, nor even of its impact. The citation rate is a *partial indicator* of the impact of a scientific publication: that is, a variable determined partly by (a) the impact of the paper on the advance of scientific knowledge, but also influenced by (b) other factors, including various social and political pressures such as the communication practices (for example, the reading and referencing habits of different individuals in different institutions, countries, and research areas), the emphasis placed on numbers of citations for obtaining promotion, tenure, or grants, [25] and the existing visibility of authors, their previous work, and their employing institution. As with numbers of publications, it cannot be assumed that the effects of (b) are relatively insignificant compared to those of (a), [26] nor that (b) comprises a set of essentially random influences, the effects of which cancel out in analyses of large aggregations of scientists or for extended time periods. The "other factors" that make up (b) are largely social and political rather than purely "scientific", and, while some of their effects on citation rates may be random, others can be expected to vary in a systematic way between individual, or groups of, scientists occupying different cognitive and social locations. The relative importance of (a) and (b) on citation rates can only be established empirically – for example by looking to see whether there are systematic variations in referencing behaviour between different groups of scientists. [27]

Once the claim for citation rates is restricted to that of being a *partial* indicator of the scientific *impact* (rather than a *measure* of the *quality* or importance) of research publications, then some of the methodological problems associated with citations that have been detailed above are considerably diminished. First, an important pioneering paper may be little cited initially, but this would suggest that such a publication also has little *impact* at first until its importance is eventually recognised, so the number of citations is still correlated with impact. The small number of early citations is therefore a reflection of the structure and organization of the scientific community and its communication system, rather than a consequence of any inherent weakness in citation analysis per se (cf. Lawani [28, p. 29]). Indeed, the use of citation rates to identify those pioneering papers whose importance is not immediately recognised may well be an essential first step in ascertaining *why* recognition of their importance is not always readily forthcoming. Secondly, a controversial, but low-quality paper, or even a "mistaken" paper, may create a large impact, stimulating in its wake further research aimed at improving upon, or refuting it. If such a paper generates a large number of citations, this can be regarded as a reflection of its impact, rather than of its quality (cf. Cole and Cole [9, p. 25]), and hence of how much it stimulates and therefore contributes to scientific progress. Thirdly, if citation rates are taken to reflect the impact of a paper rather than its quality or importance, the fact that, in cases where there has been complete integration of certain basic ideas, the original publications currently receive few, if any, citations, can be interpreted as implying that

[25] Kaplan [24, p. 183] warned that the use of citation analysis as a tool for evaluating science could lead to changes in citation practices, and in the course of our interviews with astronomers, several mentioned their belief that some individuals and groups, particularly in the United States, had already begun to exploit the system through informal mutual-citing arrangements – often termed "citation circles".

[26] Although they discuss the possible effects of various "distorting" factors on citation counts, Cole and Cole [8, p. 369] assume that the incidence of such effects is rare and their magnitude insignificant, thus enabling them to claim that "citations generally represent an authentic indicator of influence". (It should be noted that this represents a weakening from their earlier claim that the number of citations represents the *quality* of a paper – see Cole and Cole [7, p. 379].)

[27] Kaplan [24, pp. 182–3] raises some of the questions that need to be addressed in such empirical studies. For example, "what is the influence of the organisational context and of one's colleagues? Are there informal norms about citing the works of colleagues, especially of superiors, so long as there is some remote connection between their work and the work being reported?...Do citation practices reflect significant elements of the normative and value systems of scientists?" These questions have largely gone unanswered since then, although some recent work on the classification of citations by context has been carried out by Chubin and Moitra [5], and Moravcsik and Murugesan (e.g. [39,40]) – work that may eventually provide answers to such questions.

those original publications have little direct impact now. For example, Einstein's 1905 papers are still regarded as being of supreme importance, yet they have a low impact today and elicit few citations because hardly anyone actually reads the papers. Before they reached such a stage of integration into scientific knowledge, these papers had a far higher impact and would presumably have received a very high number of citations compared with other contemporary publications. Since that time the impact of these papers has diminished, while the importance of the ideas contained in them has remained high and has been transferred through second, third, and successive generations of papers, each possibly citing the "parent" generation of papers but not the "grandparent" and previous antecedent papers (cf. Margolis [31, p. 1215]). Hence, there seems to be no objection to using citation rates as a partial indicator of "impact" provided that we recognise that this is not synonymous with "importance", a quality which may be transferred through successive generations of papers in such a way that citations do not reveal the "founding father" paper.

The fourth problem, that of the high-quality paper in a small unpopular field which gains less citations than those in large active fields, has already been dealt with – the lower number of citations being interpreted as representing a smaller impact on the advance of scientific knowledge (in the sense of having implications for the work of a smaller number of researchers). Similarly, the fact that certain kinds of research papers (or papers written by more "visible" scientists) tend to receive more citations than average can be interpreted as indicating that such papers, whatever their quality, do create a wider impact than average. For example, a paper describing a minor improvement to an established method may constitute a relatively small or low-level contribution to scientific knowledge, but when integrated with other contributions it can have a big impact on scientific activities (and hence on subsequent scientific progress), altering the procedure involved in some standard and widely used technique; and this will probably be reflected in its citations, at least until it is integrated into the body of scientific knowledge. [28]

The problem of self-citation is a little more complex in that it is not immediately clear whether, when using citation rates as a partial indicator of impact, one needs to exclude self-citations or not. It can, for example, be argued that individuals or groups of scientists who engage in a high level of self-citation do so because their previous work has had a large impact on their current research. However, at the same time, it must be recognised that in the case of self-citations, the relative effect of (b), the "other factors", on the citation rate may very well be greater than in the case of "normal" citations. Perhaps the best solution is to examine empirically the effect of including both self-citations and in-house citations (citations to the work of colleagues within the centre) in order to see whether the effect is significant or not, and this is attempted in the empirical study described below.

7. Peer evaluation

The method of assessing contributions to scientific knowledge apparently most favoured by scientists is that of peer evaluation. [29] However, as with publication counts and citation analysis, peer evaluation does not yield a simple measure of scientific progress. The difficulty centres on the subjective nature of this method: peer evaluation is based on individual scientists' *perceptions* of contributions by others to scientific progress, perceptions arrived at through a complicated series of intellectual and social processes, mediated by factors other than the quality, importance, or impact

[28] There is, however, a problem with citations in that some standard techniques, even after having become fully accepted, still do generate citations to the original papers (as in the protein-determination techniques referred to in note 19, while others (for example, aperture-synthesis in radio astronomy) do not. The reasons for such differences in citation patterns need further empirical investigation.

[29] Out of the 69 scientists interviewed in this study, 76% saw peer evaluation as the best method, and 19% saw it as equal best with either publication counts or citation analysis. Fairly similar results are obtained by Chan [3] in his study of university faculty staff and administrators on the relative importance of various output indicators in evaluating the effectiveness of academic research. Peer evaluation and "articles published in prestigious journals" are seen as the most valid indicators, although it is not absolutely clear in this study whether the latter refers merely to *numbers* of publications, or whether it was interpreted by the respondents in the survey as including some assessment of the quality or influence of those papers as well.

of the research under evaluation. This gives rise to three main sets of problems in using peer evaluation, and we need to be aware of these if we are to use it as another partial indicator of scientific progress.

First, and most importantly, we must recognise the effect of political and social pressures within the scientific community on the way certain scientists will evaluate the worth of their peers. Modern science is highly competitive, with groups vying for prestige and funding. Research groups, particularly those in areas of science characterized by rapid scientific growth, complete fiercely for the recognition and acceptance of their ideas and theories, and peer review can reflect the pressures associated with this rivalry. Scientists who are asked to evaluate research may be influenced by the possible implications of the peer evaluation both for their own futures and for those of their colleagues. Of course, one could try to reduce the effects of this problem by asking every peer even remotely affected by the particular scientific contribution under assessment for his or her evaluation, and then aggregating or averaging the responses; but given that this is seldom practical, it is generally only possible to ask a representative selection of peers – and this in turn raises all those statistical problems associated with sampling. One can at best only attempt to ensure that representatives from all the important groups in the field are involved in the assessment in order to minimise these problems.

Representative sampling also helps reduce the effects of a second problem with peer evaluation, namely that of allowing for the diversity in the cognitive locations of the reviewers. [30] There is inevitably a tendency for scientists to evaluate the worth of a scientific contribution in terms of their own research interests and activities. This problem is compounded by the fact that, although peer evaluation is supposedly based on the published output of scientists, some reviewers have not read all the relevant publications, with the result that their evaluations may be based more on informal "coffee-room" discussions or on the *reputation* of scientists (in particular, their reputation for performing at conferences) than on their actual contributions to scientific progress (cf. Smith and Fiedler [50, p. 226]). One way to overcome this problem is to include in the sample of reviewers only eminent scientists with a wide knowledge of the field concerned, but this can introduce other problems (such as that of excluding the views of less well established, but often no less relevant, schools of scientific thought).

Finally, we should mention a problem associated with all attempts to gauge opinion – the propensity of people to conform to conventionally accepted patterns of belief. Although perhaps privately holding the work of certain other well-known scientists in low esteem, they may publicly express different views (even if they know that their assessments are to be treated confidentially). Such conformist behaviour need not always be the result of a conscious decision, one example being the "halo effect" by which work may be evaluated more highly because of its association with a successful group or a prestigious university. Furthermore, particularly in those cases where evaluators do not possess all the knowledge needed to make a balanced judgement, yet do not want to withdraw because an invitation to participate in peer evaluation is often regarded as an indication of status within the scientific community, there can be a tendency to base the evaluations on the *recognition* accorded to work, for example in the form of prizes or medals, rather than on the work itself. There is then the additional problem of taking account of the time-lag between the making of the contribution to science, and its subsequent recognition.

Because of these difficulties, peer evaluation cannot claim to provide an incontrovertible measure of scientific progress. As with the other measures we have considered, it is at best a partial indicator of scientific progress [31] – a variable influenced partly by the size of the contribution to

[30] Cf. the discussion by Mitroff and Chubin [36, p. 199] of the need to examine the cognitive styles of reviewers and reviewees in order to make sense of the results of peer review.

[31] Hence, one would not expect the results of peer evaluations to be altogether independent of other quantitative indicators such as publications and citations. The former is linked to the latter in that peer evaluation is based partly on journal literature, and some aspects of journal literature (for example, the acceptance of papers for publication, and the citing of others in a paper) may in turn be partly determined by peer evaluation (cf. Moravcsik [38]). In the few studies where peer evaluation is combined with other indicators of scientific progress, the various methods of evaluation are found to coincide fairly well (cf. *ibid.*).

scientific progress and partly by other factors, the relative importance of which cannot be assumed to be insignificant.

8. Other possible indicators of scientific progress

Two further indicators of scientific progress are the number of "discoveries" or other major contributions to the advance of knowledge, and the formal recognition afforded to scientists by the scientific community. In the case of "discoveries", the crucial questions centre on *who* should identify the "discoveries", and using *what criteria.* The use of the notion of "discoveries" therefore involves problems similar to those encountered with peer evaluation. It would be most unsatisfactory, for example, merely to use text-books or histories of science for identifying discoveries, since one would be relying on the authors' unspecified criteria for judging which contributions to knowledge are sufficiently influential to merit mention as "discoveries", and which are not. Moreover, such judgements reflect *prevailing* perceptions of influence, which can be substantially different from those held at the time the particular contribution to knowledge was made. A contribution that is initially highly influential may subsequently come to be judged as "incorrect" and be omitted from reconstructions found in "official histories", even though it may have stimulated much important research at the time. Alternatively, one might use review articles for identifying discoveries on the grounds that, because the time-delay is shorter, the problem of the reconstruction of history should be less severe. Such a procedure would in effect involve looking at the references cited in review articles, thus bringing one face-to-face with the problems of citations mentioned earlier. It is true that the citation practices in review articles *may* be less wayward or self-centred, and less subject to the effects of the "other factors" described above, than citation practices in research publications (perhaps because a more comprehensive literature-search generally precedes the writing of a review, or because only the "better" scientists are chosen to perform this task). However, this can only be investigated empirically – it cannot be assumed.

The other indicator of scientific progress we should consider is the recognition accorded to scientists through the awarding of various honours such as medals, prizes, invitations to lecture at major conferences, or election to a national academy of science. The allocation of such recognition involves a prior process of peer evaluation, and is, as such, based not only on scientists' perceptions of contributions to knowledge, but also on numerous other factors, the effects of which may vary considerably. For example, Crane [11, p. 710] claims to find that affiliation to a prestigious university is more likely to lead to recognition for a scientist than high productivity. Such a result leads to doubts about the adequacy of recognition as even a partial indicator of contributions to scientific progress, unless preceded by a thorough analysis of the social structure of the scientific community and of its mechanisms for allocating recognition.

9. A methodology for assessing contributions to scientific knowledge

In preceding sections we have seen how all quantitative "measures of science" are at best partial indicators, influenced by a network of interrelated factors of which the size of contribution to scientific progress is but one. Nevertheless, given that all decisions on the allocation of resources between different areas of science must involve some assessment of the respective merits of the various areas, we would argue that selective and careful use of these indicators (despite the associated methodological and conceptual problems) is far better than none at all (cf. Moravcsik [38]). Previous use of these indicators, in the field that has come to be known as "scientometrics", has met with a barrage of criticism, in particular from scientists. The purpose of the previous sections has been to show why much of this criticism is justified. For example, assessments of contributions to science based on numbers of citations can be misleading or meaningless when little or no thought has been given to the question of what the numbers actually indicate. Nonetheless, the fact that indicators of science have in the past been applied in a simplistic or irresponsible manner does not mean that they cannot be usefully employed at all. On the contrary, having analysed the shortcomings and conceptual difficulties associated with each of them, we are now in a better position to see how they might be used to achieve more reliable assessments of contributions to scientific knowledge.

Given the absence of any adequate single measure of scientific progress, the only way forward would appear to be through the combination of *several* partial indicators. However, since each partial indicator is influenced to a greater or lesser extent by numerous other factors in addition to the magnitude of the contribution to science involved, one must attempt to "control for" these other factors by choosing the groups of scientists to be compared in such a way that one can examine the size of the effects of the "other factors" on each of the partial indicators. Only in those cases where convergent results are obtained can it be assumed that the influence of the other factors has been kept small (i.e. the matching of the groups has been largely successful), and that the indicators therefore provide a reasonable estimate of the contribution to scientific progress made by different research groups. The extent to which convergent results can be obtained is examined in the empirical study described here.

As mentioned before, the study focusses not on individual scientists [32] but on research centres, because these absorb a significant proportion of basic research funds. With this unit of analysis, several of the problems with indicators discussed earlier are diminished. In particular, the influence on the indicators of the size of the contribution to scientific knowledge is less likely to be swamped by the influence of those "other factors" whose effects are statistically random, while it is also easier to match groups than individuals [33] in such a way that one controls for some of the "other factors" whose effects are systematic. Fig. 2 summarizes the main problems with the various indicators and how we have attempted to minimize their effects in our study.

In selecting research centres for comparison, what is ideally required is that there should be two or more groups, working in the same specialty over a similar time-period, publishing in the same journals, supported with a roughly similar level of resources, and situated in a similar institutional context. For example, a high-energy physics group like that using the Rutherford Laboratory might be compared with that for Daresbury Laboratory (and both with similar centres overseas). Similarly, the radio astronomy groups at Cambridge, Jodrell Bank, and a number of overseas institutions can be compared in terms or their contributions to scientific progress. It is with radio astronomy that this study is concerned.

Even when dealing with groups of scientists rather than individuals, however, no "matching" is perfect. Thus, while Cambridge and Jodrell Bank are both engaged in radio astronomy, they tend to concentrate on different research areas and problems within this specialty. This is partly the consequence of the different nature of the research equipment at each centre; Cambridge has concentrated on interferometers (for short-baseline interferometry), while Jodrell Bank has placed more emphasis on "big dishes" (used either individually, or together for long-baseline interferometry). It is therefore essential to include other centres in this study to ascertain whether this divergence in emphasis leads to systematic differences in the effects on the partial indicators. For this purpose, the Netherlands Foundation for Radio Astronomy (NFRA) group based on the Westerbork interferometer and the receiver at Dwingeloo, and the Max-Planck-Institut für Radioastronomie (MPI) group based on the Bonn "big dish", were chosen. [34] These latter two centres are essentially national facilities (as opposed to the university facilities at Cambridge and Jodrell Bank), and are both supported at several times the level of their British counterparts. Nevertheless, given the similarities in the size and scale of activity of the

[32] The great majority of previous attempts to assess contributions to scientific knowledge have concentrated on individuals rather than research groups, even though very little scientific research is now carried out by isolated individuals. The few authors who have compared *groups* of scientists include Westbrook [54], Larabi [26,27], Matheson [33], and Andrews [1].

[33] In the case of individuals, it is unlikely that one will find more than a very small number whose (fairly narrow) research interests coincide or even overlap reasonably completely; with groups, however, a wider range of research interests is covered and it should be possible to find several for which there is a substantial degree of overlap.

[34] Inevitably, this matching of centres is still not perfect. The Bonn big dish is more precise than the one at Jodrell Bank, and is therefore used for work at shorter wavelengths and for more spectroscopy. In the case of the Dutch and the two main Cambridge interferometers, the latter are used primarily for continuum work, while the former carries out both continuum and spectral-line work. Nonetheless, the degree of comparability between the centres, while not complete, is sufficient for the analysis that follows.

Partial indicator based on	Problem	How effects may be minimized
(A) Publication counts	(1) Each publication does not make an equal contribution to scientific knowledge	Use citations to indicate average impact of a group's publications, and to identify very highly cited papers
	(2) Variation of publication rates with specialty and institutional context	Choose matched groups producing similar types of papers within a single specialty
(B) Citation analysis	(1) Technical limitations with *Science Citation Index:*	
	(a) first-author only listed	Not a problem for research *groups*
	(b) variations in names (c) authors with identical names (d) clerical errors	Check manually
	(e) incomplete coverage of journals	Not a serious problem for "Big Science"
	(2) Variation of citation rate during lifetime of a paper – unrecognised advances on the one hand, and integration of basic ideas on the other (3) Critical citations (4) "Halo effect" citations	Not a problem if citations are regarded as an indicator of impact, rather than quality or importance
	(5) Variation of citation rate with type of paper and specialty	Choose matched groups producing similar types of papers within a single specialty
	(6) Self-citation and "in-house" citation (SC and IHC)	Check empirically and adjust results if the incidence of SC or IHC varies between groups
(C) Peer evaluation	(1) Perceived implication of results for own centre and competitors may affect evaluation (2) Individuals evaluate scientific contributions in relation to their own (very different) cognitive and social locations. (3) "Conformist" assessments (e.g. "halo effect") accentuated by lack of knowledge on contributions of different centres	(1) Use a complete sample, or a large representative sample (25% or more) (2) Use verbal rather than written survey so can press evaluator if a divergence between expressed opinions and actual views is suspected (3) Assure evaluators of confidentiality (4) Check for systematic variations between different groups of evaluators

Use only indicators that yield convergent results

Fig. 2. Main problems with the various partial indicators of scientific progress and details of how their effects may be minimized.

Dutch and German groups, their operation as national facilities, and their use of the same journals for publication, it should be possible to see whether carrying out research on an interferometer rather than a big dish leads to systematic differences in numbers of publications or citation-rates. If there are such systematic differences, and if they are consistent with the results based on peer evaluation, then one can perhaps conclude that these reflect actual differences in the level of scientific progress achieved by each centre rather than merely being a consequence of the "other factors" affecting the publication and citation indicators.

The main partial indicators of scientific progress used in this study are publications, citations, and peer evaluation. "Recognition" is not employed because most prizes and honours are predominantly national, [35] and so, although it could be used for comparisons between Cambridge and Jodrell Bank, it is not appropriate for comparisons between centres in different countries. There are few truly international honours, the Nobel Prize being almost the only exception, and the numbers are certainly too small for systematic use in a comparison such as this. [36] There are similar problems with an indicator based on the notion of "discovery". The number of major discoveries is rather small for this to be used in a systematic way, while broadening the category of "discoveries" may involve some arbitrariness as to what should be included as a "discovery", and what should not. [37] However, some attempt to assess relative numbers of discoveries is made by adopting highly cited papers as an indicator.

10. Input measures

Before considering each of the partial indicators of scientific progress, we must first look at the scale of scientific activity at each centre; there are appreciable differences in the sizes of the four groups, and it would therefore be misleading to compare the scientific progress made by the smallest group with that made by the largest without taking account of their respective scales of activity. There are a number of input measures that can be used to define this scale of activity, one of the most important being the number of scientists actually engaged in research. The 1978 figures [38] for this are set out in table 1. It can be seen that, in terms of researchers (excluding students), Jodrell Bank is approximately 50% larger than Cambridge, the Dutch group over twice as large, and the German group over three times the size. If research students are included, however, these dif-

Table 1
Number of astronomy researchers in 1978

	Cambridge	Jodrell Bank	NFRA	MPI
Number of astronomy researchers [a] (excluding students)	17	25	~ 40	~ 60
Number of research students [b]	16	14	~ 10	~ 15
Total number of astronomy researchers (staff and students)	33	39	~ 50	~ 75

[a] This is the number of astronomers who regularly use the observation facilities of the group and who publish in astronomical journals. It excludes short-term visiting users of the radio telescopes, but it does include a number of scientists in each group whose primary duties involve the development or maintenance of instrumental and computer facilities, but who nevertheless devote a significant amount of their time to astronomical research.

[b] This excludes M.Sc. students at Jodrell Bank, first-year postgraduates at Cambridge, "Doctorandus" students in Holland, and Diploma students in Germany; in other words, we are comparing like with like, as far as possible.

[35] For example, it is doubtful whether one can equate a Fellowship of the Royal Society in Britain with membership of the National Academy of Sciences in the United States because of differences in the percentages of practising scientists selected in each country, in the criteria used in selection, and so on.

[36] It is probably significant, however, that the Cambridge group, which, as we shall see later, appears to have contributed most to scientific progress, includes two scientists who shared the Nobel Prize for Physics in 1974.

[37] This is reflected in the responses made by astronomers to a question on what have been the major achievements of each of the groups. There is, for example, a high degree of consensus that the first observations of pulsars represented a major discovery, but far less consensus over various lower-level achievements.

[38] Ideally, the corresponding figures for earlier years should also be included, but these have proved far harder to obtain. Nonetheless, one can make certain general comments about the trends over time. At the start of the ten-year period upon which we have focussed, the Dutch and German groups were in their infancy. They grew rapidly over the first five years or so, but over the next five years (up to 1978), the increases in staff numbers and running costs have been much slower, approaching the modest growth rates experienced by the British groups over the ten years. Hence, during those last five years, the *ratios* of the scale of activity at the four centres have not changed very significantly.

ferences are to some extent reduced.

A second factor influencing the level of scientific activity is the amount of time radio astronomers must devote to teaching activities. At both Cambridge and Jodrell Bank, faculty staff report that they spend on average about 30% of their time on teaching and administration of undergraduate and M.Sc courses, while Dutch and particularly German astronomers have far lighter teaching loads. Clearly the more teaching that is carried out, the less time that remains for research, and one way of considering this is as a reduction in the effective number of astronomy researchers using each of the centres. The respective adjustments for each centre are shown in table 2. It can be seen that this accentuates the differences in scale of research between the two British centres and their two continental counterparts. [39]

A third factor to be considered is the annual running costs [40] for the group. The 1978 values for these are presented in table 3, together with various figures on the costs per researcher depending on whether students are included or not, and on whether differences in teaching duties are allowed for. The figures on total running costs show similar or slightly larger differences in scale than the figures in tables 1 and 2, the larger differences being attributable to the higher salaries in Holland and Germany (see the figures on costs per researcher). In addition, the major capital investment costs (that is, the cost of a new telescope, or of extensive refurbishing of an existing instrument) are summarized in table 4. It should be noted that, if these are spread over the likely life-time of a radio telescope (20 years, for example), they are small in comparison with the running costs, even after allowing for the effects of inflation. Consequently, it is the running costs that largely determine the ratios of the total costs for the centres.

A final factor that affects the level of scientific activity in a research group is the number of support staff (i.e. non-astronomers), and the figures for this are shown in the second row of table 5, along with values for the ratio of support staff to astronomers. It can be seen that Cambridge has the smallest number of support staff, Jodrell Bank about 15% more, the Dutch group double the Cambridge number, and the German group is nearly three times better off in this respect. However, the ratios of support staff to astronomers are broadly similar for all the centres.

To summarize: all the input indicators suggest that the scale of scientific activity is greatest for the German group, which is well over twice as large as the smallest group, at Cambridge. Jodrell

Table 2
Effective number of astronomy researchers in 1978 after allowing for teaching duties [a]

	Cambridge	Jodrell Bank	NFRA	MPI
Numbers of researchers involved in teaching	14	17	~ 30	~ 20
Percentage of time spent on teachning [b]	33%	31%	~ 10%	~ 10%
Number of researchers effectively "lost" to research	5	5	~ 3	~ 2
Effective number of researchers (excluding research students)	12	20	~ 37	~ 58
Effective number of researchers (including research students)	28	34	~ 47	~ 73

[a] This covers time spent on teaching undergraduate, M.Sc., Dutch "Doctorandus", and German Diploma students.
[b] Individual astronomers were asked how much of their time was devoted to teaching, and the results were averaged for each centre to give these figures.

[39] Astronomers also spend a certain part of their time in research-related administration both within their centre and outside (on national and international committees, etc.). In addition, some researchers devote much effort to developing new equipment. The time taken up with such duties has not been separated from that devoted to research, for the reason that these duties are an essential and integral part of modern, large-scale research. While research centres overall spend a roughly similar fraction of their time on instrumental tasks and administration, this fraction obviously varies cyclically as a centre passes through the stages of designing a major research facility, construction, exploitation, designing a new facility, and so on. As a result, in the 10-year period under consideration, Jodrell Bank claim that they devoted rather more of their efforts to instrument-building than their rivals.

[40] Because the *total* expenditure at a centre can obviously fluctuate appreciably from one year to the next depending on whether a major new piece of equipment is being built, in calculating the running costs for each centre, we have excluded major items of capital expenditure (of greater than £100,000). However, 'running costs', as defined here, do cover *all* other expenditure, including the cost of purchasing less expensive and more routine items of equipment.

Table 3
Annual running costs for 1978 [a] (£ Sterling)

	Cambridge	Jodrell Bank	NFRA	MPI
Annual running costs [b]	£0.6 M	£0.9 M	£2.6 M	£3.7 M [d]
Annual cost per astronomy researcher (excluding students)	£35 k	£35 k	£65 k	£60 k
Annual cost per astronomy researcher (including students)	£20 k	£25 k	£50 k	£50 k
Annual cost per effective researcher [e] (excluding students)	£50 k	£45 k	£70 k	£65 k
Annual cost per effective researcher (including students)	£20 k	£25 k	£55 k	£50 k

[a] This is the estimated expenditure in the calendar year 1978 after excluding the major items of capital expenditure shown in table 4. Where the financial year does not coincide with the calendar year, an appropriate average of the expenditure in financial years 1977/78 and 1978/79 has been taken. Because of this, and because of the different accounting procedures at each of the centres, and the consequent difficulties in deciding which resources to include or exclude, the figures quoted here are accurate to only 5 or 10%. This is however, quite sufficient for our purposes. It should be noted that in an earlier draft of this paper, a rather wider definition of 'running costs' was used, which is why the figures shown in this table are slightly smaller than those quoted previously.

[b] The 1978 exchange rates of 3.8 German Marks and 4.2 Dutch Guilders to the British Pound are used in these calculations.

[c] This includes the cost of the salaries for approximately 30 astronomers and 10 Ph.D. students from the universities of Leiden, Groningen, and Utrecht, who regularly use the NFRA research facilities, together with the salaries of 15 support staff.

[d] This includes the cost of the salaries for approximately 15 researchers and students (and of their support staff) at the universities of Bonn, Hamburg, Bochum, and Tübingen, who regularly use the MPI telescope.

[e] I.e. after excluding time spent teaching.

Bank appears to be about 20% larger than Cambridge, and the Dutch group over 50% larger. In terms of running costs, the differences in scale between the groups are apparently more pronounced, probably because of the higher salaries in continental Europe, with the Dutch group costing about three times as much and the German group over four times as much as each of the British groups.

Table 4
Details of major items of capital expenditure at each radio astronomy observatory up to 1978

Centre	Details	Completion date	Approximate capital cost
Cambridge	1-mile interferometer	1964	£0.5 M
	5 km interferometer	1972	£2.1 M
Jodrell Bank	Mark I 250-ft telescope	1957	£0.7 M
	Mark II 125×83-ft telescope	1964	£0.3 M
	Mark III 125×83-ft tele scope	1967	£0.1 M
	Mark IA conversion	1971	£0.5 M
	Mark VA design costs [b]	1974	£0.7 M
	Multi-telescope radio-linked interferometer (MTRLI) – Knockin telescope	1977	£1.9 M
NFRA	1.5 km interferometer	1970	£3.0 M [a]
	Addition of 2 movable telescopes to 1.5 km interferometer	1975	£0.9 M
MPI	100 m telescope	1971	£4.0 M [a]

[a] In converting the Dutch and German costs into Sterling, we have used the yearly average exchange rates quoted in National Institute of Economic and Social Research [42].

[b] This telescope, which was intended to be a national, rather than a solely Jodrell Bank facility, was not in fact built. Details of the project and of the decision not to proceed with it can be found in a report by the Public Accounts Committee [48].

11. Publications

The first of the output indicators we shall consider is numbers of publications. These are shown in table 6, together with the aggregate and average for the ten years up to 1978 for each centre. [41] Only articles in properly refereed, scientific journals and in published conference proceedings (i.e. publications generally recognised by scientists as being contributions to science) are included. Articles providing popularised accounts of scientific research, unpublished work, preprints, [42] and

[41] These figures are based on the complete publication lists provided for us by each of the centres, and they should therefore be reasonably accurate. If there are errors, they are unlikely to be more than about 5%.

[42] In radio astronomy, nearly all preprints are eventually published; they are therefore excluded in order to avoid "double-counting".

Table 5
Number of support staff in 1978 [a]

	Cambridge	Jodrell Bank	NFRA	MPI
Total number of staff at the centre	90	104 [d]	160 [b]	240 [c]
Number of support staff (excluding astronomy researchers)	57	65	110	165
Number of support staff per researcher (including students)	1.7	1.7	2.2	2.2
Number of support staff per effective researcher (excluding students)	4.7	3.3	3.0	3.0
Number of support staff per effective researcher (including students)	2.0	1.9	2.3	2.3

[a] As in the previous tables, the figures may be slightly inaccurate because of the difficulties in estimating what to include and what to exclude in cases where resources are shared with other research groups. For example, the staff of the Mullard Radio Astronomy Observatory at Cambridge is 84, but there are in addition some employees, such as cleaners, telephone switchboard operators, and librarians who are shared with the Cavendish Laboratory, Cambridge University, bringing the total to about 90. The overall accuracy of the figures in the table should, however, be sufficient for our purposes.

[b] This includes 30 university astronomers and 10 students at Leiden, Groningen, and Utrecht, who carry out their research primarily on the NFRA facilities, together with an estimated 15 support staff (secretaries, cleaners, etc.) at these universities.

[c] This includes 15 astronomers and students at Bonn, Hamburg, Bochum, and Tübingen, together with an estimated 8 support staff at these universities.

[d] This excludes 16 staff formally employed to provide services to Jodrell Bank visitors (in the Concourse Building, restaurant, and Arboretum), a number of whom also provide some support services to the radio astronomers.

laboratory reports are all excluded.

From these figures, one can see that the level of scientific production of the Cambridge group grew fairly steadily until the late 1970s when it levelled off. The slight "dip" in output in the early 1970s coincides with the period during which a large proportion of the group's effort was devoted to bringing the new 5-km interferometer into operation (see table 4). In the case of Jodrell Bank, the level of production grew until the end of the 1950s, during which time it was consistently higher than at Cambridge; after this, it dipped in the mid-1960s (when the Mark II and Mark III telescopes and multi-channel digital spectrometer were built), and has again declined more recently as a major effort has been put into the new multi-telescope radio-linked interferometer (see table 4). However, even after allowing for these temporary fluctuations, the average level of scientific production of Cambridge has still been considerably higher than that of the Jodrell Bank group over the ten years up to 1978.

As we have already seen, the numbers of publications is but a partial indicator of scientific pro-

Table 6
Numbers of publications

	Cambridge	Jodrell Bank	NFRA	MPI
Publications in				
1946	1	1		
1947	1	5		
1948	6	11		
1949	2	5		
1950	5	8		
1951	6	13		
1952	9	14		
1953	5	13		
1954	7	20		
1955	13	18		
1956	10	19		
1957	18	20		
1958	10	7		
1959	6	24		
1960	18	24		
1961	21	20		
1962	19	20		
1963	17	22		
1964	12	20		
1965	22	17		
1966	30	11		
1967	43	15		
1968	48	22	15	
1969	40	21	16	16
1970	32	22	22	45
1971	32	21	41	41
1972	43	27	37	35
1973	33	23	65	57
1974	47	16	55	47
1975	40	21	66	65
1976	40	16	57	78
1977	46	18	92	87
1978	45	12	91	79
1969–78 aggregate	398	197	542	550
1969–78 average	40	20	54	55

gress. While the difference between the publication rates for Cambridge and Jodrell Bank *may* indicate a difference in scientific progress, it may also be accounted for by other factors. For example, it may be related to the different types of radio telescope and research interest at the two centres. Alternatively, there may be substantially different publication practices in the two groups with the result that Jodrell Bank publications on average make a more substantial impact in terms of their contribution to scientific knowledge. The latter possibility is examined in the next section on citations, but the former can be considered here by looking at the output of the Dutch and German groups.

During the ten years up to 1978, the production of papers by both the Dutch and German groups has increased rapidly, finally reaching a level about twice that of the Cambridge group and more than four times that of Jodrell Bank. This high volume of production for the two continental groups can be partly explained in terms of their higher funding and staffing levels. If one allows for this (see table 7), it appears that the production of the Cambridge and the Dutch groups has been relative high – certainly higher than that of the German and Jodrell Bank groups (even after allowing for the latter's recent diversion of effort into instrument-building – see note c to table 7). Whether the smaller number of publications for the latter is compensated for by each publication representing, on average, a greater contribution to scientific knowledge is considered in the next section.

Table 7
Numbers of publications relative to staffing and funding levels [a] in 1978 [b]

	Cambridge	Jodrell Bank [c]	NFRA	MPI
No. of papers per astronomy researcher (excluding research students)	2.6	0.5 (0.8)	2.3	1.3
No. of papers per astronomy researcher (including research students)	1.4	0.3 (0.5)	1.8	1.1
No. of papers per effective researcher [d] (excluding research students)	3.8	0.6 (1.0)	2.5	1.4
No. of papers per effective researcher (including research students)	1.6	0.4 (0.6)	1.9	1.1
Approximate cost per paper	£15 k	£75 k (45 k)	£30 k	£45 k

[a] The likely errors in these figures are probably somewhere between 5 and 10% (see note a to table 3 and note a to table 5.

[b] These are the figures for a single year only (1978), and it is is conceivable that they might be untypical of the decade 1969–78. However, for the reasons mentioned in note 38, it is unlikely that figures for the four centres would show a very different pattern in previous years, at least once the two continental centres began operating at full strength.

[c] The figures in parentheses are those obtained for Jodrell Bank if 1978's publication rate of 12 papers is assumed to be a temporary "dip", and the figure of 20 papers (the ten-year average) is substituted as being a more typical value.

[d] I.e. after excluding time spent on teaching (see table 2).

12. Citations

Figures on the numbers of citations to *all* previous research work by each centre have been calculated, using their complete publications lists and the *Science Citation Index*. These are detailed in table 8. The number of citations to Cambridge work has increased by over 150% during the period 1966–1978, with a slight pause to this growth evident in the 1974 figures. Table 9, giving figures for citations to work published over the previous four years, suggests that this temporary fall can probably be attributed to the diversion of effort at the beginning of the 1970s away from research and into the commissioning of the new 5-km interferometer. [43] Both the Dutch and German groups also show a rapid growth in the numbers of citations to their work, reaching, in 1978, a total similar in magnitude to that for Cambridge. In the

[43] The effects of this concentration on instrument-building show up even more clearly in table 11, which indicates that the citation rate per Cambridge publication dropped markedly in the early 1970s. During this period many of the staff were involved in building, testing, and commissioning their new telescope, and, while the publication rate of the group as a whole fell only slightly (largely kept up by some highly productive PhD students), the citation rate for each paper declined dramatically.

Table 8
Citations to all previous work [a]

	Cambridge	Jodrell Bank	NFRA	MPI
1966 citations	500	420	–	–
1970 citations	860	410		
1971 citations			210	
1972 citations				320
1974 citations	810	490	580	550
1978 citations	1380	470	1360	1030
1978 citations to 1969–78 work only	1120	340	1340	1030

[a] It is difficult to attach any estimate to the likely errors in these figures. If they were of the order of $\sqrt{n}$, this would correspond to percentage errors of between 3 and 7%.

Table 10
Rates of self-citation (SC) and "in-house" citation (IHC) in 1978

	Cambridge	Jodrell Bank	NFRA	MPI
Number of citations analysed in sample [a]	217	200	200	200
No. of SCs	9	4	8	9
No. of IHCs	32	28	36	30
Percentage of SCs	4%	2%	4%	5%
Percentage of IHCs	15%	14%	18%	15%
Total % of SCs and IHCs (likely error)	19% (±3%)	16% (±3%)	22% (±3%)	20% (±3%)

[a] This sample of 1978 citations was chosen on a random basis.

case of Jodrell Bank, however, the number of citations to previous work has scarcely changed during the 12 years studied. Similarly, table 9 shows that the most recent work by the Cambridge, Dutch, and German groups has all been far more highly cited than that by Jodrell Bank.

In the earlier theoretical discussion, it was argued that citation rates should be regarded as a partial indicator of scientific progress. This being so, we must now examine the extent to which the lower citation rate for Jodrell Bank reflects a smaller overall contribution to scientific progress than those of the other groups, or is merely a product of certain "other factors". But what might such "other factors" include? Can, for example, the low citation rate be attributed to weaker links with the international community of radio astronomers, less of whom therefore cite Jodrell Bank work? Scientists in both Holland and Germany commented, in interviews, that *both* the Cambridge and the Jodrell Bank groups tend to be somewhat "closed" in their attitudes towards outsiders, but they claimed that this tendency is more pronounced at Cambridge than at Jodrell Bank. Hence, it would seem difficult to attribute the latter's low citation rate to this cause. Can the differences in numbers of citations for each group be attributed to differing rates of self-citation and "in-house" citation [44] amongst the centres? A sample of about 200 1978 citations for each centre was examined, and table 10 shows that there is little significant difference between the four centres in the frequency with which they cite their own group's work.

Yet another possibility is that the total number of citations for the work of a group is very largely determined by its publication rate. We can examine whether this is a satisfactory explanation by looking at the average number of citations to each paper. Table 11 shows the figures for the citations to work published in the preceding four years, divided by the number of publications produced in that time, i.e. the average number of citations per publication (CPP). From table 11, it can be seen that the Cambridge group, while having a far lower overall output of papers than the two European groups, nevertheless achieves a significantly higher number of citations per publication. This suggests that the number of citations is not just a function of numbers of publications.

Table 9
Citations to work published in the previous four years. [a]

	Cambridge	Jodrell Bank	NFRA	MPI
1970 citations	540	200		
1971 citations			210	
1972 citations				320
1974 citations	330	220	480	340
1978 citations	550	190	780	610

[a] See note a to table 8.

[44] I.e. citations by a scientist to the work of colleagues in his or her group.

Table 11
Citations per publication (CPP) for work published in the previous four years. [a]

	Cambridge	Jodrell Bank	NFRA	MPI
CPP in 1970	3.3	2.5		
CPP in 1971			2.2	
CPP in 1972				2.3
CPP in 1974	2.1	2.5	2.4	1.9
CPP in 1978	3.2	2.8	2.4	2.0

[a] The likely errors in these figures probably range between 3 and 7% (see note a to table 8).

There would, therefore, seem to be some justification for arguing that, for the radio astronomy centres in this study, citation rates do provide a useful partial indicator of scientific progress, In the case of total citation rates, the figures certainly *suggest* that the Cambridge, Dutch, and German groups have made rather more impact on the advance of knowledge than Jodrell Bank in the period 1969–1978, with Cambridge having done particularly well in relation to its size. The figures for citations per publication again suggest that, of the four centres. Cambridge papers have generally had the largest impact (with the exception of those at the start of the 1970s – see note 43); in this respect, Jodrell Bank seems to have done slightly better than the Dutch and German groups, although the *total* impact of all its papers is apparently still much smaller because of its lower publication rate. However, until we take account of the results of peer evaluation, we cannot determine with any confidence whether such differences in citation rates are significant or not, and whether figures on citations say anything about the overall contribution of each centre.

Table 12
Highly cited papers published between 1969 and 1978 [a]

	Cambridge	Jodrell Bank	NFRA	MPI
No. of papers cited 15 or more times in 1 year [b]	12 *	1	6 †	1 *
No. of times these highly cited papers received 15 or more citations in 1 year	23	1	13	1
No. of papers cited 20 of more times in 1 year	4 *	0	3 †	0 *
No. of papers cited 12 or more times in 1 year	19 *	3	7 †	2 *

* Excludes papers based on work carried out elsewhere by astronomers before moving to the centre.
† Includes one review paper, and one paper describing the new synthesis telescope.
[a] Rigorous efforts have been made to ensure that these figures are as accurate as possible. It is unlikely that any of the values in the first two rows is wrong by more than 1 or 2 units. Errors of this magnitude would have no effect on the conclusions drawn in the text concerning the distribution of highly cited papers between the centres.
[b] Out of the total of approximately 1700 publications produced by the four centres between 1969 and 1978, only 20 (i.e. about 1.2%) have been cited 15 or more times in any one year.

Before considering the peer-evaluation results, however, there is one further argument that could be levelled against the preceding citation analysis, and which needs to be scrutinised. It is argued by some (e.g. Cole and Cole [8]) that it is not the great mass of publications (each constituing but a minor increment to the sum of human knowledge) which contributes most to scientific progress, but a small number of key papers, each of which has a very great impact on the advance of knowledge. If this is the case, then the figures for total numbers of citations, and for citations per publication, may not reveal which centres have managed to produce such key papers and which have not. However, we can establish which centres have produced the most highly cited papers and, although (as was seen earlier) there are problems in using citations for comparing individual papers, this should give some indication as to the distribution of key papers between the four centres. [45] Figures for approxi-

[45] One might suppose the number of highly cited papers to be at least a partial indicator of the number of "discoveries" made by a research group, and there is indeed some evidence to support this supposition from the peer-evaluation interviews described later. In these, astonomers were asked to identify the principal "discoveries" or significant contributions to scientific knowledge made by each of the research groups over the ten years up to 1978. There was a high degree of consensus among the interviewees on at least the major contributions, and these generally coincided with the most highly cited papers. Moreover, at one of the two centres with very few highly cited papers, many of the astronomers freely admitted that the centre had been responsible for no major discovery during the ten-year period.

mately the top 1% most cited papers (cited 15 or more times in any one year) are shown in table 12. It can be seen that between 1969 and 1978 the Cambridge group produced twelve such papers, and the Dutch group half that number, while the two other groups managed only one each. Some of these publications were highly cited only for a short time, after which their impact appears to have diminished rapidly; the second row of table 12 takes account of the variation between these short-lived papers and those that have been highly cited over many years by showing the number of times these papers managed to achieve 15 or more citations in a year. Again a similar pattern emerges, although the differences between the four groups are, if anything, further accentuated. (Lest it be imagined that this distribution of highly cited papers is in some way the result of choosing an arbitrary value of 15 citations a year as the dividing line between highly cited and less highly cited papers, table 12 also includes the numbers of papers cited 20 times or more, and 12 times or more, in each case yielding a similar distribution between the four centres.)

13. Peer evaluation

In a series of approximately 70 interviews with virtually the entire scientific staffs of the two British groups, and with about one-third of the Dutch and German groups (including all the most senior astronomers), scientists were asked to identify the main contributions to scientific knowledge made over the period 1969–1978 by their own centre and by eight other major radio astronomy centres. [46] This group included all the radio observatories considered by the astronomers in our survey as being the foremost centers in this field. They were then requested to rank these centres in order according to the magnitude of their contributions to scientific knowledge made over the ten-year period. [47] We subsequently investigated the consistency of the evaluations produced within each centre, and the consistency between centres. In particular, we looked for systematic differences between groups, such as a tendency to over-rate one's own centre.

The average rankings (on a scale between 1 and 9 [48] because there are nine centres) obtained at each centre are shown in table 13. A number of comments can be made about the results. First, there is a fairly high degree of consistency in the results obtained both *within* each centre, and *between* the centres. Indeed, if one averages the results for all four centres, the great majority (31 out of 36 [49]) agree with the average rankings for all four centres to within one unit or less.

Secondly, there is *some* tendency for scientists to over-rate the scientific progress made by their own group, but while this is significant in two of the four centres, [50] it is not as marked as might have been expected, and does not appear to be sufficiently large to cast doubts about the overall significance of the peer-evaluation results.

Thirdly, it is possible to detect a slight influence of certain particularistic factors on these rankings. For example, Cambridge ratio astronomers tended to rank interferometer centres slightly more highly,

[46] Besides the four centres already discussed, these included the American radio telescopes at Arecibo, Caltech, and the National Radio Astronomy Observatory (NRAO), the French centre at Nançay, and the Australian big dish at Parkes. We have, as yet, not looked at publication and citation indicators for five of the nine centres, but hope to do so in the future to see how the results correlate with the results of our peer evaluation. This would then constitute a study of a large part of the world radio astronomy community, and the results would be of relevance not only to science policy, but also to the sociology of science.

[47] As might be expected, there was a certain reluctance to do this, generally on the grounds that, because each centre is involved in slightly different work, it is misleading to rank them on a single scale. Nevertheless, most respondents (88%) did agree to take part in this ranking exercise; some felt able to rank each centre in order from 1 to 9, while others preferred to group the centres into categories of "first class", "second class", "third class", etc.

[48] If two centres were ranked first-equal, they were given the average ranking of 1.5. Similarly, if three were ranked equal first, they were given the ranking of 2 (the average of 1, 2, and 3); and so on.

[49] Four of the remaining five differed by only 1.1 units, and the remaining one, where the discrepancy was 2.2, was a self-ranking – i.e. the ranking by one centre of its own position on the scale.

[50] In addition, the senior MPI staff tended to slightly over-rate themselves, but this was cancelled out by the more junior staff who, if anything, tended slightly to under-rate the scientific progress of the group. This difference between the evaluations of senior and junior staff at the German centre may be related to a certain dissatisfaction amongst the latter about the organization and orientation of the group's research.

Table 13
Rankings of nine radio astronomy centres [a] obtained by peer evaluation (PE)

	PE at Cambridge ($n = 11$) *	PE at Jodrell Bank ($n = 19$) *	PE at NFRA ($n = 13$) *	PE at MPI ($n = 18$) *	Average rankings for all four centres	Average rankings excluding self-rankings
Arecibo	7.8	7.3	7.0	6.7	7.2 (±0.4)	7.2 (±0.4)
Caltech	4.6	6.0	6.7	5.3	5.7 (±0.8)	5.7 (±0.8)
Cambridge	2.0	3.1	1.9	1.6	2.2 (±0.6)	2.2 (±0.6)
Jodrell Bank	5.7	3.8	6.1	6.1	5.4 (±1.0)	6.0 (±0.2)
MPI	6.3	4.8	4.3	5.3	5.2 (±0.7)	5.1 (±0.8)
Nançay	8.7	8.9	8.9	8.9	8.8 (±0.1)	8.8 (±0.1)
NFRA	2.5	3.3	1.6	2.7	2.5 (±0.6)	2.8 (±0.3)
NRAO	3.2	4.3	2.8	2.6	3.2 (±0.7)	3.2 (±0.7)
Parkes	4.0	3.6	5.7	5.6	4.7 (±0.9)	4.7 (±0.9)

* n = sample size.
[a] Ranked in terms of the magnitude of their contributions to scientific knowledge between 1969 and 1978. The figures in parentheses in the final two columns indicate the root-mean-square variations between the rankings given by the different centres.

and some of the big-dish centres less highly, than average, suggesting that the type of equipment and type of research at each centre may affect the peer evaluations. Similarly, radio astronomers who had previously worked at one of the other centres tended to rank that centre more highly than average. For example, some of the American astronomers working in the Dutch and German groups ranked the American radio astronomy centres slightly more highly than average. These examples illustrate one of the potential problems with the use of peer evaluation as an indicator of scientific progress – that scientists may evaluate more highly those centres whose work is most familiar to them. However, our results suggest that, while this may be true for some individuals, the effects of these particularistic factors are barely significant when using fairly large samples of peer reviewers. [51]

The high degree of consistency in the peer-evaluation results, together with the relatively small effects of self-rating and other particularistic factors, gives some support for the belief that peer evaluation, at least when carried out with a large representative sample of scientists from carefully matched groups, can yield a reasonably good indication of the scientific progress made by research groups. Now, given that some confidence can be placed in the peer-evaluation results, what do the results actually show? The average rankings of the nine groups are shown in fig. 3, [52] these averages having been obtained after excluding self-rankings on the ground that these may sometimes produce a small but nevertheless significant effect on the results. From this figure it can be seen that the Cambridge and Dutch groups, together with the National Radio Astronomy Observatory (NRAO) in the United States, are clearly perceived as having been world-leaders in radio astronomy over the period 1969–1978. They are followed by a group of four other observatories, including Jodrell Bank and Bonn, with the former apparently in the lower half of this group. The two remaining groups were seen as having achieved significantly less

[51] The American astronomers working in Holland and Germany ranked two of the three American observatories slightly more highly than average, but the differences are scarcely significant. Arecibo was ranked 6.3 (±1.3) compared with the average figure of 7.2 (±0.4) and NRAO was ranked 2.8 (±1.3) compared with the average of 3.2 (±0.7). The third American centre at Caltech was ranked below average (at 6.4 (±1.8) compared with the average of 5.7 (±0.8)), but again the difference is probably not significant. Because of the relatively small size of this systematic deviation, it seems doubtful whether the overall peer-evaluation results would be substantially altered by including radio astronomers from countries other than those covered here. Nevertheless, it would be interesting to include the evaluations of French, American, and Australian radio astronomers, and this may be attempted in the future.

[52] It must be emphasised that this is *not* a graph; it is merely a pictorial representation of the numerical results in the right-hand column of table 13. The nine centres have been arranged in the order in which they were ranked solely to bring out the differences perceived by radio astronomers between the top three observatories, the second group of four, and the remaining two centres.

scientific progress over this period. [53] Of the four centres with which we are principally concerned, it can be seen that the Cambridge group is regarded as having been the most successful of the four over the ten years up to 1978, while the Jodrell Bank and Bonn groups have apparently achieved rather less scientific progress (or that at least is the perception of most radio astronomers).

It is important to note that fig. 3 presents a static picture; it does not show changes in the relative positions of the centres over time. Besides ranking the observatories in terms of their performance over the ten-year period, we did ask radio astronomers to identify any changes in position *during* that period. From their responses it is evident that, at the start of the ten years, the Cambridge group was clearly perceived as being *the* world leader in radio astronomy, while the big dishes at Parkes and Jodrell Bank had also proved highly successful over the preceding period. However, in subsequent years, there appears to have been a significant improvement in the relative positions of NRAO and of the two newcomers, NFRA and MPI (in line with the changes over time noted in the indicators discussed earlier). This improvement came at the expense of Parkes, Jodrell Bank, and Caltech (which suffered from the increasing concentration of American radio astronomy resources on NRAO). Cambridge's lead was challenged first by the Dutch group when their new interferometer began operating, and is now about to face a severe challenge from NRAO's new Very Large Array. And perhaps in the near future, the MTRLI (see table 4), which was under construction in the 5 years up to 1978, may restore Jodrell Bank to something closer to its original position in the early 1960s. Certainly the number of publications produced by Jodrell Bank has increased significantly since this work was carried out (to 20 in 1979), and the extent to which this represents an increased contribution to scientific knowledge relative to that made by other radio observatories would be an interesting subject for future study.

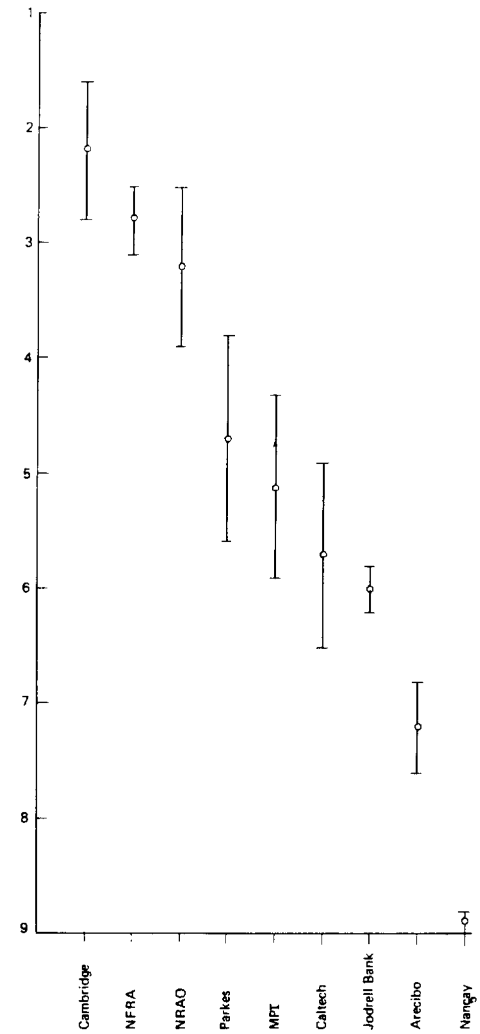

Fig. 3. Average rankings of nine radio astronomy groups for 1969–78 (excluding self-rankings). (The error-bars indicate the root-mean-square variations between the rankings given by the different centres.)

14. Discussion and conclusions

We began this paper by discussing the need for comprehensive, basic-science policies in modern industrial societies, where large amounts of state expenditure are now concentrated on a small number of "Big Science" centres. Perhaps the key

[53] In the course of the interviews, several other radio astronomy groups were mentioned as having achieved significant scientific progress. These included the groups at Bologna in Italy, Penticton in Canada, and Bell Telephone Laboratories in the United States. Some of these groups, if they had been included in our list of observatories, might have ranked as highly as the middle group of Parkes, Caltech, and Jodrell Bank, perhaps pushing some of these centres down the rankings. None, however, were seen as being in the same rank as the top three observatories.

problem in the past has been the lack of an adequate methodology for evaluating "success" in basic research. In the theoretical discussion in the earlier part of the paper, after analysing the limitations of such measures as publication counts, citation frequencies, "discoveries", and peer evaluation, we put forward a framework based on a set of partial indicators for comparing the scientific contributions made by research groups working in the same specialty. This was then applied to evaluate the performance of the two principal radio observatories in Britain – with NFRA and Bonn brought in as the major control groups. There are just two questions remaining to be addressed in this final section. First, to what extent are we justified in using the methodology described here to compare and to differentiate the scientific performance of basic research groups? And secondly, what policy implications can be drawn from the results obtained in this way?

The most important thing to note is that there is a high degree of consistency between the results obtained with four of the partial indicators. Numbers of highly cited papers and peer evaluation both suggest that the *total* contributions of the Cambridge and Dutch groups have been relatively large; and, if we allow for the differences in scale of research activity between the centres, the two partial indicators based on publications per researcher and citations per paper [54] suggest that the research productivity of the Cambridge group has been the highest of the four centres. This *convergence* in the results suggests [55] the selection of the control groups for comparison with the two British centres has been such that the effects of the "other factors" on the partial indicators have been kept relatively small, and hence that some reliability can be placed in assessments based on these four indicators. So what do these four indicators suggest for each of the radio astronomy centres? The results of this study suggest the following conclusions for each of the centres:

(1) The Cambridge radio astronomers are perceived by their peers as having achieved as much scientific progress over the decade 1969–78 as any other group in the world. Although there are larger and better funded groups, this disadvantage is offset partly by the Cambridge scientists achieving a high publication rate per researcher (especially if one allows for the time taken up with undergraduate teaching), and partly by producing papers which, on average, have more impact on the advance of scientific knowledge. They have also produced a comparatively large number of highly cited papers, [56] a result which is consistent with the judgement of their peers that they have made several major discoveries.

(2) The Dutch radio astronomers have achieved as much, or almost as much, as their Cambridge colleagues in terms of contributions to scientific knowledge. As at Cambridge, each researcher on average produces a large number of publications, so that, overall, the group produces more papers than Cambridge. However, the peer-evaluation results imply that the total Dutch contribution to scientific knowledge is no greater than for Cambridge, which in turn suggests that each Dutch publication does not constitute such a substantial contribution to knowledge, a conclusion apparently supported by the figures on citations per publication. Moreover, the group's research, although funded more generously, has resulted in significantly fewer, highly cited papers and, according to their peers, fewer discoveries than the Cambridge work.

(3) The large number of radio astronomers at Bonn ensures that, despite a somewhat lower publication rate per person, the group as a whole produces a large number of scientific

[54] The two other indicators – total numbers of publications and citations – do not yield results consistent with these four, and are therefore apparently not very reliable indicators of scientific progress. For example, both suggest that the German group has achieved as much as the Cambridge and Dutch groups. The reason is that numbers of publications reflect the level of scientific activity and production of a research centre rather than the scientific progress it is making, while a large total number of citations can be generated by publishing a large number of papers, even if each makes little impact on the advance of scientific knowledge.

[55] Convergence does not, of course, "prove" that the indicators are "reliable". There is still a possibility that, despite all the precautions taken, they are "unreliable" but linked together (see note 31), which is why conclusions can only be *suggested* rather than unreservedly drawn.

[56] 3.0% of Cambridge papers achieved 15 or more citations in a year, compared with 0.9% for the Dutch group, 0.5% for Jodrell Bank, and 0.2% for Bonn.

papers. The overall scientific progress made by the group is not, however, as great as for the Cambridge and Dutch groups, at least as judged by peer evaluation. This can be interpreted as implying that each German publication, an average, has a rather smaller impact on the advance of knowledge, a conclusion again in line with the figures for citations per publication. In addition, the Bonn group has not produced very many highly cited papers, nor many "discoveries" in the opinion of their peers.

(4) In the case of Jodrell Bank, the publication rate for the group as a whole is relatively low compared to that for the three other observatories, even after allowing for time taken up with undergraduate teaching and for the diversion of effort into the new interferometer project. This somewhat lower level of production is partly offset by producing papers of higher impact than those of the Dutch and Germans. Overall, however, the contributions to scientific knowledge by Cambridge and Dutch astronomers are judged by their peers to be significantly greater than those of their Jodrell Bank colleagues, and those of the German group may be slightly larger too, although, given that in terms of effective number of researchers (see table 2) the German group is over twice that of the Jodrell Bank team, any difference between the scientific progress made by these two groups may be mainly the result of this difference in scale of activity. There is certainly no difference evident in the numbers of highly cited papers – neither group has performed very well in this respect relative to NFRA and, in particular, to Cambridge.

It is not our task in this paper to make specific recommendations about individual centres. That is the job of science policy administrators responsible for the funding and organization of basic research. The decisions they reach will be dependent upon other factors such as the available funding, their analysis of the reasons why certain centres have achieved less than others, and their judgements as to whether it is possible to devise measures to improve the performance of the less successful centres. For example, one factor which has apparently influenced the scientific performance of radio observatories is the type of instrumentation employed. The indicators based on numbers of highly cited papers and on peer evaluation both suggest that the interferometer centres have been more successful than the big-dish centres over the period in question. What remains unclear is the extent to which this is a "real" effect (in the sense that interferometers have actually contributed more to scientific progress than big dishes), and the extent to which it is an artefact of the indicators employed. However, in view of the fact that the sample of radio astronomers used in the peer-evaluation process was not obviously biased towards interferometers (we actually interviewed rather more big-dish than interferometer astronomers), the peer-evaluation results do suggest that the effect is more real than artificial, and that the period between 1968 and 1978 should consequently be regarded as the era of the interferometer in radio astronomy.

Another factor that can influence the scientific performance of a research centre is the fraction of time its research staff devote to instrument-building, a fraction that may vary cyclically as new telescopes are brought into operation. To a certain extent, the relative performance of Jodrell Bank over the period 1969–78 may be linked to their concentration on instrument-building during much of this period.

However, we shall leave a discussion of the reasons for the differences in scientific performance of the various centres to a later paper, and return to the main objective of this paper, which has been to address two questions: (a) can basic research be assessed? (b) more specifically, can significant differences in the research performance of radio astronomy centres be identified? We would contend that the evidence presented in the paper is sufficient to justify a positive answer to both these questions. However, it should perhaps be emphasised that we have been concerned with the assessment of *past* performance, while policy-makers are clearly interested in the *future* performance of research groups. Nevertheless, past performance, although by no means the only factor, is one of the best indicators of future performance, particularly for Big Science where it is virtually impossible to set up a new centre or a major research programme overnight. We therefore see it as crucial that those who have to judge the respective claims for funds by different research groups, do so on the basis of full information on the recent performance of those groups. Although further work is obviously needed, we believe that this paper goes some way in demonstrating how that information can be obtained. It establishes a methodology that, while not able to compare *directly* research centres in different specialties, is able to identify those centres that are amongst the international leaders in their own specialties, and which, if other factors are equal or indeterminate, should be given a relatively high priority in their claims for research funds.

References

[1] F.M. Andrews (ed.) *Scientific Productivity: The Effectiveness of Research Groups in Six Countries* Cambridge University Press, Cambridge, 1979).

[2] A.E. Bayer and J. Folger, Some correlates of a citation measure of productivity in science, *Sociology of Education* 39 (1966) 381–90.

[3] J.L. Chan, Organisational consensus regarding the relative importance of research output indicators, *Accounting Review* 53 (1978) 309–23.

[4] H. Chang and D. Dieks, The Dutch output of publications in physics, *Research Policy* 5 (1976) 380–96.

[5] D.E. Chubin and S.D. Moitra, Content analysis of references: Adjunct or alternative to citation counting, *Social Studies of Science* 5 (1975) 423–41.

[6] K.E. Clark, *America's Psychologists: A Survey of a Growing Profession* (American Psychological Association, Washington, DC, 1957).

[7] S. Cole and J.R. Cole, Scientific output and recognition: A study in the operation of the reward system in science, *American Sociological Review* 32 (1967) 377–90.

[8] J.R. Cole and S. Cole, The Ortega hypothesis, *Science* 178 (1972) 368–75.

[9] J.R. Cole and S. Cole, *Social Stratification in Science* (University of Chicago Press, Chicago, 1973).

[10] J.R. Cole and S. Cole, Citation analysis, *Science* 183 (1974) 32–33.

[11] D. Crane, Scientists at major and minor universities: A study of productivity and recognition, *American Sociological Review* 30 (1965) 699–714.

[12] D.L. Croom, Dangers in the use of the Science Citation Index, *Nature* 227 (1970) 1173.

[13] D. Dieks and H. Chang, Differences in impact of scientific publications: Some indices derived from a citation analysis, *Social Studies of Science* 6 (1976) 247–67.

[14] C. Freeman, Measurement of output of research and experimental development, *UNESCO Statistical Reports and Studies no. 16,* ST/S/16, Com 69/XVI-16A (1969).

[15] E. Garfield, Citation indexes in sociological and historial research, *American Documentation* 14, no. 4 (1963) 29–31.

[16] E. Garfield, Citation indexing for studying science, *Nature* 227 (1970) 669–71.

[17] E. Garfield, Citation analysis as a tool in journal evaluation, *Science* 178 (1972) 471–79.

[18] E. Garfield, What scientific journals can tell us about scientific journals, *IEEE Transactions on Professional Communication* PC-16, no. 4 (1973) 200–02.

[19] E. Garfield, Citation and distinction, *Nature* 242 (1973) 485.

[20] G.N. Gilbert and S. Woolgar, The quantitative study of science; An examination of the literature, *Science Studies* 4 (1974) 279–94.

[21] H. Inhaber and K. Przednowek, Quality of research and the Nobel Prizes, *Social Studies of Science* 6 (1976) 33–50.

[22] J. Irvine and B.R. Martin, The economic effects of Big Science: The case of radio astronomy, *Proceedings of the International Colloquium on the Economic Effects of Space and Other Advanced Technologies, Strasbourg, 28–30 April 1980* (Ref. ESA SP-151, Paris), 103–16 (1980).

[23] N.C. Janke, Abuses of citation indexing, *Science* 156 (1967) 892.

[24] N. Kaplan, The norms of citation behaviour: Prolegomena to the footnote, *American Documentation* 16 (1965) 179–84.

[25] T.S. Kuhn, *The Structure of Scientific Revolutions* (University of Chicago Press, Chicago, 1970).

[26] J. Larabi, Mesure de l'efficacité des laboratoires de recherche fondamentale sélectionnes par le Centre National d'Etudes Spatiales, *Revue Francaise d'Informatique de Recherche Operationelle* 3 (1969) 103–12.

[27] J. Larabi, Note sur l'efficacité des laboratoires de recherche fondamentale sélectionnes par le CNES, *Le Progres Scientifique* 137 (1970) 4–18.

[28] S.M. Lawani, Citation analysis and the quality of scientific productivity, *Bioscience* 27, no. 1 (1977) 26–31.

[29] D. Lindsey, Production and citation measures in the sociology of science: The problem of multiple authorship, *Social Studies of Science* 10 (1980) 145–62.

[30] J. Maddox, Is the literature worth keeping?, *Bulletin of the Atomic Scientists* 19, no. 9 (1963) 14–16.

[31] J. Margolis, Citation indexing and evaluation of scientific papers, *Science* 155 (1967) 1213–19.

[32] B.R. Martin, Cognitive and social locations: Their role in the processes of discovery and evaluation within science (Department of Liberal Studies in Science, Manchester University), mimeo (1977).

[33] A.J. Matheson, Centres of chemical excellence?, *Chemistry in Britain* 8 (1972) 207–10.

[34] A.J. Meadows and J.G. O'Connor, Bibliographical statistics as a guide to growth points in science, *Science Studies* 1 (1971) 95–99.

[35] J.D. McGervey, Citation analysis, *Science* 183 (1974) 28–31.

[36] I.I. Mitroff, and D.E. Chubin, Peer review at the NSF: A dialectical policy analysis, *Social Studies of Science* 9 (1979) 199–232.

[37] M.J. Moravcsik, Measures of scientific growth, *Research Policy* 2 (1973) 266–75.

[38] M.J. Moravcsik, A progress report on the quantification of science, *Journal of Scientific and Industrial Research* (India) 36 (1977) 195.

[39] M.J. Moravcsik and P. Murugesan, Some results on the function and quality of citations, *Social Studies of Science* 5 (1975) 86–92.

[40] M.J. Morvavcsik and P. Murugesan, Citation patterns in scientific revolutions, *Scientometrics* 1 (1979) 161.

[41] C.R. Myers, Journal citations and scientific eminence in contemporary psychology, *American Psychologist* 25 (1970) 1041–48.

[42] National Institute of Economic and Social Research, *National Institute Economic Review* 86 (1978) table 25.

[43] Is your lab well cited? *Nature* 227 (1970) 219.

[44] More games with numbers, *Nature* 228 (1970) 698–99.

[45] University of Houston expels professor, *Nature* 279 (1979) 278.

[46] A.L. Porter, Citation analysis: Queries and caveats, *Social Studies of Science* 7 (1977) 257–67.

[47] Price, D. de S., *Little Science, Big Science* (Columbia University Press, New York, 1963).

[48] Public Accounts Committee, Expenditure on design for proposed Mark VA radio telescope, *Fifth Report from the Committee of Public Accounts,* Session 1975–76, HC 556 (HMSO, 1976 London), pp. xxxi–iv, and 269–78.

[49] E. Shils (ed.), *Criteria for Scientific Development: Public Policy and National Goals* (MIT Press, Cambridge, Mass., 1968).

[50] R. Smith and F.E. Fiedler, The measurement of scholarly work: A critical review of the literature, *Educational Record* (1971) 225–32.

[51] D. Sullivan, D.H. White and E.J. Barboni, The state of a science: Indicators in the specialty of weak interactions, *Social Studies of Science* 7 (1977) 167–200.

[52] N. Wade, Citation analysis: A new tool for science administrators, *Science* 188 (1975) 429–32.

[53] A.M. Weinberg, Criteria for scientific choice, *Minerva* 1 (1963) 159–71.

[54] J.H. Westbrook, Identifying significant research, *Science* 132 (1960) 1229–34.

[55] R.D. Whitley, Communication nets in science. Status and citation patterns in animal physiology, *Sociological Review* 17 (1969) 219–33.