

DISCUSSION

H. M. COLLINS*

WHEN DO SCIENTISTS PREFER TO VARY THEIR EXPERIMENTS?

Introduction

REPLICATION of experiments is a subject much neglected in the philosophical literature. Franklin and Howson's paper¹ is especially to be welcomed for its appeal to scientific practice in support of the analytical argument. However they give too little attention to the representativeness of the cases they discuss and to the means by which their data are generated. Insufficient attention to these matters leads to a false impression of the strength and generalisability of their argument. In fact, their argument, with some additional assumptions, can be applied to only that subset of scientific work where there is substantial agreement among the experimenters. In these cases, of course, replication of results is less crucial than in areas of disagreement.

I take Franklin and Howson's argument to amount to the following: an experiment which is *identical* to another cannot support it. This must be true, since if it were truly identical, Heraclitean in their terms, then looking at the results of the second experiment would amount to no more than looking a second time at the results of the first — for example re-reading the original experimental report. An experiment can only confirm the results of another to the extent that it is *significantly* different in one respect or another.

I am sure this argument is correct in principle. It is however *inapplicable*. In cases of *substantial disagreement or uncertainty* — cases where the question of confirmation is most important — it assumes too much knowledge on the part of experimenters. In cases of *disconfirmation* it is wrong. In nearly all cases it assumes agreement on the part of experimenters over experimental competence, and over demarcation criteria.

The difficulty arises out of the authors' treatment of 'theory' as an articulated, monolithic consensual feature of experimental design. Their unusual approach to evidence concerning the procedures of the 'physics community' may be the problem. They take these procedures from statements

*Science Studies Centre, University of Bath, Claverton Down, Bath BA2 7AY. I am grateful to David Gooding for a helpful discussion of this paper. Fieldwork discussed was supported by a grant from the Social Science Research Council.

Received 18 October 1983.

¹A. Franklin and C. Howson, 'Why do Scientists Prefer to Vary Their Experiments?', *Studies in History and Philosophy of Science* 15 (1984), 51.

in *Review of Particle Physics* and other 'leading research journals'. However it is well known that the formal journals² present only a partial view of scientific work; as long ago as 1963 Medawar wrote an article called '*Is the Scientific Paper a Fraud?*'³ Understanding has advanced considerably since then and in particular we now know the research paper presents an unrealistically ordered, logical and consensual account of scientific practice.

Variation unnecessary for confirmation in poorly understood areas

As Franklin and Howson are aware, the idea of an *identical* experiment is an abstraction. As they point out, even to begin to think of 'Heraclitean repetitions' involves assumptions. For example, time-invariance is assumed when it is said that the *same* experiment was repeated later. But in the early stages of experimental work on a new phenomenon little is self-consciously assumed; the first act of an experimenter who has done an easy experiment which produces an unusual result will be to do it again, immediately. In these circumstances the experimenter is checking to see that the first result was not an 'anomaly' or other unspecified 'mistake'. Data runs cannot be treated as additive, with the invariance assumptions that this implies, until the experimental procedures are well established. The examples that Franklin and Howson provide, in which results are simply added, come from well established areas where questions of confirmation no longer apply.

In disputed areas, where confirmation is the issue, repetitions of any sort give support to an experimental claim. Given the in-principle validity of Franklin and Howson's major argument in a situation of perfect knowledge, how can this be? The answer is that methodological practice depends upon perceptions or assumptions about similarity and difference; there is no perfect knowledge in practice. In new areas scientists recognise that there are many unknown and unknowable factors. These make all replications potentially non-trivially different however similar they seem. Thus, identity of results of experiments is not guaranteed at the outset however Heraclitean they appear to be.

A simple example is the set of factors related to the experimenters themselves. However Heraclitean in intention, a confirmation by a known rival, or a disconfirmation by a known ally and colleague, has more power to convince than the same results produced by the opposite parties. This is easier to keep in mind if discussion is couched in terms of variations in the '*experimental system*'; this includes the socio-cognitive relationship between

²By 'formal journals' I mean journals which direct themselves toward other specialists, as opposed to news journals such as *New Scientist* or *Scientific American*.

³P. Medawar, 'Is the Scientific Paper a Fraud?', *Listener* 12 September 1963, pp. 377–8.

experimenters.⁴ In controversial areas the identity of the experimenter is widely perceived as being a crucial variable even though this is never apparent in the formal literature. *Ad hominem* reasoning is more important in these areas than experimental design or articulated theory.

So far it has been argued that in ill-understood new areas scientists do not feel the need to vary their experiments deliberately in order to confirm their results. Any sort of repetition is perceived as adequately non-identical since ignorance of the nature of the experimental conditions makes it impossible to guarantee results. Interestingly, the value of intended Heraclitean experiments diminishes as an area becomes better understood, that is, just as the need for confirmation diminishes.

It has been argued that there are circumstances in which variation in experimental design is not highly valued. There are other circumstances in which it is *similarity* of design which is actively preferred. This happens where disconfirmation of another's result is the aim of an experiment. For example, the purpose may be to reveal some unspecified inadequacy in the work of the original experimenter or his apparatus (experimental system). If critics' experiments appear to differ significantly from the original design then these differences can be taken to account for the failures to find confirming evidence. Thus critics' negative results must be produced by experiments that appear to be *the same* as the originator's if they are to be convincing.

The history of the failure to detect high-fluxes of gravitational radiation in the early 1970s illustrates this point. In 1972, by which time a number of negative results had been recorded, Joseph Weber, the original claimant, remarked as follows:

Well, I think it is very unfortunate because I did these experiments and I published all relevant information on the technology, and it seemed to me that one other person should repeat my experiments with my technology, and then having done it as well as I could do it they should do it better . . . It is an international disgrace that the experiment hasn't been repeated by anyone with that sensitivity.⁵

As late as 1975, even Weber's critics were able to find fault with every one of the critical experiments done to date, except for the one experiment done in Germany which set out to be a 'carbon-copy' of Weber's. (Critics found fault with one another's experiments, rather than their own.⁶)

⁴For a discussion of this problem see R. Rosenthal, *Experimenter Effects in Behavioural Research* (New York: Irvington, 1976), and the interesting panel discussion of this work in *The Behavioural and Brain Sciences* 3 (1978), 377.

⁵Quoted in H. M. Collins, 'The Seven Sexes: a study in the sociology of a phenomenon, or the replication of experiments in physics', *Sociology* 9 (1975), 205.

⁶See H. M. Collins, 'Son of Seven Sexes: the social destruction of a physical phenomenon', *Social Studies of Science* 11 (1981), 33.

Scientists Disagree About Similarity and Difference

There are then, circumstances in which variation is not valued highly and others in which it is negatively valued. A more fundamental point concerns the difficulty of reaching agreement about what constitutes similarity and difference in experimental systems. As Franklin and Howson point out, similarity and difference vary as a function of background theory. They conclude, rather weakly,

The fact that scientists, being fallible, are occasionally wrong in their judgements about significant differences between experiments does not argue against our model presented above. It shows only that differences may be hard to find.

The problem is more important than this conclusion suggests.

In areas where confirmation is important, agreement about theory and therefore agreement about similarity and difference is *co-extensive* with agreement about the meaning of the outcome of the sequence of tests! This is because an understanding of the experiments depends upon an understanding of the new phenomenon under examination, and vice versa. If we imagine a case (such as is discussed in references cited in notes 5 and 6) where two sets of experiments give conflicting results, yet the corresponding experimenters believe they each have the better experiments, then one set of experiments (it could be both) must be incompletely understood. Agreement about which set is to be counted as properly understood is co-extensive with agreement about which results are the correct ones — that is, it is co-extensive with agreement about the existence and nature of the phenomena in question.

Before these agreements are reached similarity and difference will be a hotly disputed matter. It is easy to see the potential for dispute in the circumstances described in the previous section of this paper. At the same time, during such periods the 'artful', skilful, non-orderly features of scientific experiment, usually disguised by reporting conventions, will become relatively visible. Scientists will be aware of their incomplete understanding of their own experimental systems even as they argue for the superiority (or essential similarity — see previous section) of their own work. The following four quotations from four different scientists involved in the gravitational radiation controversy illustrate this tension; they can be read in conjunction with the quotation given above.

You can pick up a good text book and it will tell you how to build a gravity wave detector . . . At least based on the theory that we have now . . . Basically it's all nineteenth century technology and could have been done a hundred years ago except for some odds and ends.

The thing that really puzzles me is that . . . everybody . . . is just doing carbon copies. That's the really disappointing thing.

. . . it's very difficult to make a carbon copy. You can make a near one, but if it turns out that what's critical is the way he glued his transducers, and he forgets to tell you that the technician always puts a copy of Physical Review on top of them for weight, well, it could make all the difference.

Inevitably in an experiment like this there are going to be lots of negative results when people first go on the air because the effect is that small, any small difference in the apparatus can make a big difference in the observations . . . I mean when you build an experiment there are lots of things about experiments that are not communicated in articles and so on. There are so-called standard techniques but those techniques, it may be necessary to do them in a certain way.⁷

We can see then, that not only is obvious variation in experimental design not always valued, and not only is it sometimes negatively valued, but that agreement on what counts as variation and difference is hard to reach in contentious areas.

Conclusion

Franklin and Howson are right in that, other things being equal, scientists do prefer to vary their experiments. In consensual areas of science their theory might apply, but the consensus must be substantial. For example, there must be sufficient agreement over the nature of the phenomenon in question to allow for there to be agreement over what is to count as a well performed experiment.⁸ In areas where confirmation is difficult their analysis clearly breaks down. However, as a matter of fact scientists generally *prefer* to vary their experiments even in such areas. As we have seen this preference can only be weakly related to confirmatory potential; the preference is probably much more strongly related to the system of rewards for different types of work in science.

Analysis of replication confirmation and disconfirmation may add a great deal and gain a great deal by referring to empirical studies. Franklin and Howson's formal analysis is fully applicable only to cases of perfect knowledge and approximately applicable only to cases of substantial

⁷*Op. cit.* note 4.

⁸There must also be agreement over demarcation criteria. *Too much* experimental or theoretical variation is not welcome. To take an extreme example, confirmation of some natural phenomenon by an astrologer would not be welcome! When is an experiment sufficiently *like* another to count as a confirmation?

For a fuller exploration of the possibility of an analytic theory of replication, inspired by Franklin and Howson's paper, see chapter 2 of H. M. Collins *Changing Order* (Sage, forthcoming late 1984).

agreement over theory and experimental technique — that is, the 'text book' model of science. For those who wish to use empirical studies to further their analyses there is now a body of published material available.⁹ Difficulties will be avoided if the problem is thought of in terms of *experimental systems*, which include the experimenter, and in terms of scientists' *perceptions* of similarity and difference; the latter are by no means uniform even during the same historical period, so that abstract discussion of similarity and difference leads to inapplicable models.

⁹See for example the papers by Travis, Pickering, Harvey and Pinch in *Social Studies of Science* 11 (1981) and see the footnotes to the editorial introduction by Collins in the same issue.

For reviews of a wider literature which includes many relevant articles see S. Shapin 'History of Science and its Sociological Reconstructions', *History of Science* 20 (1982), 157, the introductory sections and bibliography of *Science in Context*, S. B. Barnes and D. O. Edge (eds.) (Milton Keynes: Open University Press; Cambridge, Massachusetts: M.I.T. Press, 1982), and H. M. Collins, 'The Sociology of Scientific Knowledge: Studies of Contemporary Science', *Annual Review of Sociology* 9 (1983), 265.