Part XII

THREE VIEWS OF EXPERIMENT

Allan Franklin, Right or Wrong

Robert Ackermann

University of Massachusetts, Amherst

I regret to inform you that Allan Franklin is unable to be here because of the consequences of his collision with a truck in Boulder, Colorado several weeks ago. The three of us have decided to proceed with the symposium in his honor, even though it is now missing its fulcrum. The original point of the symposium was to have an informed discussion of two versions of atomic parity-violation experiments, versions that embody opposed philosophical conceptions of the experiments. The first conception is embodied in Andy Pickering's account in his *Constructing Quarks*, an account that is explicitly criticized by Allan Franklin's more recent discussion in his *Experiment, right or wrong* (Pickering 1984; Franklin 1990). The symposium would have brought this confrontation into focus, with Franklin's presentation of his critique followed by Pickering's rejoinder at this symposium, both of them in sufficient command of the detailed history of the atomic parity-violation experiments to allow for the possibility of a useful exchange of differences of opinion. In Franklin's absence, Pickering commands the field of relevant scientific detail here by default, and we are reduced out of courtesy to the missing position to more general issues concerning normative and constructionist accounts of experimentation. After we have spoken briefly in turn about these more general issues, we will have whatever discussion may be provoked from the floor before ending the session.

In Allan Franklin's absence, I have taken on the task, not just of chairing the symposium, but of presenting his views sufficiently so that the papers of Pickering and Lynch have a semblance of live context. Franklin's new book is a continuation of the historical studies he began in his earlier *The Neglect of Experiment* (Franklin 1986). The point of this work (in conjunction with well known studied by Latour and Woolgar, Pickering, Lynch, Hacking, Knorr-Cetina, and others) is to rescue the notion of experimentation that seems so crucial to an understanding of science from the disembodied form that it took in older empiricisms where experimentation was regarded as a mechanism for producing data regarded as factual assertions that could be used to test the truth claims of theory. As these recent and quite varied detailed studies of experimentation show, experimentation is a complicated concrete process culminating in data that are often subject to different interpretations, that is, data that are not as fact-like as the older empiricisms had assumed. Franklin is probably to be singled out on this terrain, not only for his background expertise as an experimental physicist, but for

PSA 1990, Volume 2, pp. 451-457

his normative attitude that experiments can be divided (in principle) into the good and the bad in sufficient time to effect valid discriminations between rival scientific theories. In short, Franklin's work, in the full context of these studies, has to be located in the area of that position which sees experiment as providing an environment of settled fact that selects among theoretical mutants, rather than seeing experiment as one element that is articulated or negotiated with approximately equal weight against other elements in a sort of open scientific dialogue.

It's pretty easy to see Franklin's general attitude encapsulated in the title of his new book, *Experiment, right or wrong*, and even more clearly in his Preface. The Preface recounts Richard Feynman's efforts to visit Tannu Tuva, a small Asian country well-known for its postage stamp issues, "in the right way," that is, without taking advantage of his prominence as a scientist. After several years of inquiry and preparation, Feynman died two weeks before permission arrived. Franklin's moral to this narrative is that science should proceed like Feynman did in seeking admission to Tannu Tuva, that is, "in the right way." The assumption that there is a right way, based on the epistemology of experiment to be developed in the book, is transparent. But, of course, other morals can be drawn from Feynman's quest, perhaps the most obvious being that if you search for something in the right way, it will always arrive too late. This seems to me to be the nub of any discussion of Franklin's work. Can the right way of experimentation, whatever that is, be identified sufficiently early on that the advance of science can be plotted along rational paths, in some clear sense of rationality? Whatever that sense of rationality, it apparently must contain a normative component that identifies a right way for rational scientists, against which individual deviations have to be seen, not as variants increasing the social gene pool of scientific possibilities, but as errors.

Without going into the details of the atomic parity-violation experiments, the issue between Franklin and Pickering comes down to an agreement that the scientists involved *chose to accept* certain experiments and their interpreted results, but to a disagreement as to what it means to say that these choices were reasonable. In Franklin's imagery, Pickering regards the experiments that were not accepted as mutants that were slain by a decision not to let them live, whereas Franklin regards them as mutants that died of natural causes, i.e., mutants that died because they were bad experiments, and could not be nourished in an appropriate field of data.

Let me quote Franklin quoting Pickering:

> We saw in the preceding section that in 1977 many physicists were prepared to accept the null results of the Washington and Oxford experiments and to construct new electroweak models to explain them. We also saw that by 1979 attitudes had hardened. In the wake of experiment E122, the Washington-Oxford results had come to be regarded as unreliable. In analysing this sequence, it is important to recognize that between 1977 and 1979 there had been no *intrinsic* change in the status of the Washington-Oxford experiments. No data were withdrawn, and no fatal flaws in the experimental practice of either group had been proposed. What had changed was the *context* within which the data were assessed. Crucial to this change of context were the results of experiment E122 at SLAC. In its own way E122 was just as innovatory as the Washington-Oxford experiments and its findings were, in principle, just as open to challenge. But particle physicists *chose* to accept the results of the SLAC experiment, *chose* to interpret them in terms of the standard model (rather than some alternative which might reconcile them with the atomic physics results) and therefore *chose* to regard the Washington-Oxford experiments as somehow defective in performance or interpretation (1990, p. 174).

Commenting on this passage, Franklin says:

> Though I do not dispute Pickering's contention that choice was involved in the decision to accept the Weinberg-Salam model, I disagree with him about the reasons for that choice. In Pickering's view, "The standard electroweak model unified not only the weak and electromagnetic interactions: it served also to unify practice within other diverse traditions of HEP (high-energy physics) theory and experiment ... Matched against the mighty traditions of HEP, the handful of atomic physicists at Washington and Oxford stood little chance" (Pickering, *Constructing Quarks*, pp. 301-2). In my view, the choice was a reasonable one based on convincing, if not overwhelming, experimental evidence (1990, p. 174).

Later, summarizing an intervening exposition and discussion of the experiments, Franklin says:

> My interpretation of this episode differs drastically from Pickering's. The physics community chose to accept an extremely carefully done and carefully checked experimental result that confirmed the Weinberg-Salam theory. This view is supported by Bouchiat's 1979 summary. After hearing a detailed account of the SLAC experiment by Prescott, he stated "To our opinion, this experiment gave the first truly convincing evidence for parity violation in neutral current processes ... In addition, the most plausible alternative to the W-S model, that could reconcile the original atomic physics results with the electron scattering data, was tested and found wanting. There certainly was a choice, but, as the "scientist's account" or evidence model suggests, it was made on the basis of experimental evidence. The mutants died of natural causes (1990, pp. 180-181).

So, what is under dispute comes down to this question: Does a reasonable choice of an interpreted experiment as supporting some theoretical conjecture rather than another come down to accepting that the experiment provides (normatively) reliable experimental evidence, where reliability can in some sense be objectively calculated, or does it mean that to be reasonable is to agree to abide by a consensus arising out of open negotiation concerning all of the aspects involved in some area of scientific investigation. In other words, can truth be grounded in science in any stronger way than appealing to the limits of an (admittedly fallible) scientific consensus? Can rationality be discerned in the decisions of individuals, or is it a property of a group process in scientific investigation? It's fairly obvious that this revives an old epistemological dispute about the nature of truth that recurs in the history of philosophy, and just as obvious, I think, that the problem may result from supposing that there are just two basic positions to be considered.

Franklin is determined to use Bayesian theory to develop a notion of reliable evidence that would lay down formal tracks along which a rational discussion of evidence would have to move. It's a crucial question whether his Bayesian representation proves anything, or whether it is simply a formal representation of his already existent assumptions about the relative weight of evidence. I suspect the latter. For example, a basic assumption that gets coded into Franklin's Bayesian analysis is that the validity of an experimental result is increased by independent confirmation, that is, by the same result obtained from two different experiments that are regarded as somehow equivalent. I wouldn't quarrel with the concrete examples that are given by Franklin (and Hacking) of this phenomenon; but the question is whether a general characterization of this phenomenon can be represented in Bayesian symbolism. Franklin's informal representation of the Bayesian principle goes like this:

https://doi.org/10.1086/psaprocbienmeetp.1990.2.193089 Published online by Cambridge University Press

> Thus, if we wish to know the correct time, it is better if we compare watches than if either of us looks at our own watch twice (1990, pp. 107-108).

This is hardly self-evident. For example, if I have a watch known to be very accurate, and you have an erratic watch, we may agree that two looks at my watch (to verify that it is running) can give a better estimate of the time than averaging the time shown on your watch with the time shown on my watch, particularly if they show a large divergence. Something like this happens in the relevant scientific examples, since not just any two experiments increase a consensus about validity in practice. A new experiment improving an old design completed by a scientist known to be a good experimenter may indicate an accuracy of data far superior to that offered by comparing these data to the data of an arbitrarily chosen other experimenter. It's quite obvious that scientific gossip and folklore correctly influences experimental interpretation and estimates of which experiments should be compared, but it's far from obvious that this fact can be captured in any Bayesian representation of evidential relationships between experimental results before the negotiation of gossip and discussion has reduced the field of all data to the data that count and some feeling for relationships of reliability between data derived from different experimental sources, at which point the Bayesian representation can model well enough what might have been the thinking of those who have turned out to be correct. In short, Bayesian formalism, like other formalisms, depends upon reasonable background assumptions in its use; otherwise, it can lead us far from intended goals. Curiously, Franklin's informal gloss captures this fact, congenial to constructionism, in its hypothesis that *we* wish to know the correct time (jointly), otherwise it has no application. The point I'm making does not necessarily tell in favor of constructionism. After all, the anthropological or sociological investigations of working scientific laboratories that we have were all undertaken by invasions of laboratories well known to the scientists under observation, so that they cannot be regarded as free from well canvassed distortions that can be found in such investigations. There simply is no way of objectively telling whether laboratory conversations, no matter how apparently informal, but recorded by investigators under circumstances where the investigation can't be concealed, represent what would occur in the absence of investigation, or represent instead a version of what the investigated population thinks should be its representation. Even if such investigations are enormously helpful in correcting certain kinds of idealist and normative misconceptions about the practice of science, there should be worry that these investigations may misrepresent what occurs in laboratories because the style of constructivist investigation overemphasizes the goal of cooperation in reaching agreement *within* laboratories, tends to assume that the cessation of overt disagreement means that consensus has been reached, assumes that everything can be (in principle) questioned in laboratory discussions because of an already existing view that science is going to be descriptively rational, and so forth. The problem is that laboratory talk, like all human talk, is based partly on silences that have to be *interpreted* by an investigator. Merely recording them, but not discussing them, may amount to supposing that they are not important, which is very likely not the case.

Lynch's work is much closer to constructionism as I've described it than Pickering's, but it's not at all clear that any form of constructionism needs to conflict with Franklin's form of experimental realism, provided that it would make sense to think of levels of scientific description for different purposes, in which the apparent relative disorder of high resolution analyses gives way to apparent order on lower resolution, just as relative molecular chaos may give way to regular cellular processes or individual confrontations of many kinds between individual soldiers may give way to the loss of the right flank in the analysis of a battle. Franklin's argument that the scientific community should be expected to occasionally go against the weight of evidence if the constructionist account were true is nugatory if this is taken into account. Constructionist accounts trace "weight of evidence" as it is under negotiation, in real time, until what the weight of evidence is can be subjected to consensual agreement. After that, Franklin has to be right, but his notion of weight of evidence may always come too late to catch the constructionist account opening a window on scientific irrationality. It should also be noted that the practice of following a series of related experiments on a single topic, followed by all of the approaches under consideration, may lose sight of side paths switching in and out of these sequences. Such switching can provide comparisons to changing weight of evidence in related fields that are not explicitly noted in the sequences under study, either in publications or inside scientific conversations, since one scientist affected by results in another field may have no knowledgeable audience for relevant observations. The community of conversation may only be able to communicate (without costly learning episodes) on the focus of community investigation.

Pickering's modulating position between purely constructionist description of the micro-practices of laboratories and Franklin's normative coercions can be seen in the conceptual apparatus that he brings to his discussion of experimentation. Let's turn to his interesting observation that the "scientific articulation of the real is the product of a pragmatically achieved, three-way reinforcement between material practice, instrumental modelling of the practice and modelling of the phenomenal world" (1987). This view is developed from the study of a particular sequence of experiments that Pickering analyses as though these three factors were all plastic resources equally open to adjustment until a satisfactory resonance between all three could be achieved. This is very clever, if only because by positing three equally plastic resources, the indeterminate 3-body collision problem is modelled, and the path that investigation will follow cannot be predicted. Pickering doesn't say that all three resources are equally plastic, but by calling them all "plastic resources," and noting that any of the three can be adjusted at any point, Pickering strongly suggests this reading, and the suggestion would be supported by all of those laboratory studies that tend to suggest that everything is, in principle, open to question and negotiation.

I think it would be worth exploring the existence of possible asymmetries in the plasticity of Pickering's resources, since such asymmetries might provide a clue as to how constructionism, if it's an accurate description of scientific practice, can also capture the failure of relativism that ought to turn up in an accurate description of scientific practice, without defeating relativism in advance by a notion that terms in scientific discourse correspond to items in the real world, so that there is an intrinsic metric of truth in scientific discourse. What are the three plastic resources? One (A) is the material resource of the apparatus or material experimental set-up, another (B) is the conceptual resource that explains the working of A, and the third (C) is a theoretical model or set of such models. (A, B, and C may denote appropriate sequences.) In some sense, A and B together yield data that are relevant to an evaluation of C, while C stands in turn as an evaluator of A and B, since data relevant to C must be produced. The traditional point of comparison is that between the data and C, but since the data are produced from A and B, the conceptual variability inherent in B prevents any naive experimental realism with respect to the data. This is an elegant representation of ideas to be found in Hacking (especially) and other sources, here extended by Pickering to provide a conceptual resource for discussing the development of a series of physical experiments until such a time as the problem initiating the experiments might be regarded as settled.

At the start of a sequence of experimentation, we can assume that A, B, and C are distinct resources. When such a sequence ends, typically, A and B have collapsed,

https://doi.org/10.1086/psaprocbienmeetp.1990.2.193089 Published online by Cambridge University Press

and the reliable data now being generated ends the experimental sequence with certain consequences for C. That A and B typically collapse is reflected in the fact that scientific papers find it sufficient to state B and the data, that is, to present an account of how the finally successful apparatus works. Typically, the data and C do not collapse, so that data and theory (or at least one of the theoretical models in C) remain in a tension sufficient to fuel further scientific development of a new triad (A', B', C') can be generated by a plausible or interesting variation in one of the endpoints of the original sequence. Thus Pickering's account involving three plastic resources and standard accounts are pretty much equivalent at the end of an experimental sequence. It's in the interval, as A and B are brought into resonance, that we need to concentrate our attention. Let's assume that A and B are not in synch at the start of some sequence, necessitating a change in A or B or both. Is there anything we can say about changing the apparatus as opposed to changing the theory of the apparatus? Change in either A or B can result from change in the other. But there is a difference. Some changes in A can be seen as improvements in terms of a valuation that is not sensitive to variations in B or in C. Getting an apparatus to run more smoothly or more quickly, for example, can be an obvious improvement that may (or may not) necessitate a change in B. If a change in B is required due to the relationship of A and B, it usually can be accomplished. Perhaps it has always been accomplished. On the other hand, when changes in B can be seen to be improvements, it is not always possible to change A to fit B because of some some material consideration, and sometimes changes in A that result show that the theoretical improvement was illusory. Although this does not begin to initiate a detailed analysis, there is a sense in which A appears to be a less plastic resource than B or C. To repeat, changes in A can often be seen (in real time, without waiting for accomodation by B) as *improvements*, whereas "improvements" in B don't begin to count unless A is actually altered and realizes the improvements conjectured. It's conceivable that this small asymmetry can account, ultimately, for large scale directions of scientific progress and for the objectivity and rationality of those directions.

Why isn't this possibility more widely recognized? I think the answer is that writing about experimentation automatically privileges B and C, that is, talk about experimentation, since that's what can be written down. There is the further fact that grounding rational lines of inquiry in lucky discoveries of improvement in apparatus seems embarrassing to experimenters, who might like to be granted powers of thought, and who might also crave an image of scientific rationality. Therefore, it is not all that frequent that an experimental paper freely admits that a breakthrough occurred when someone tried some "sticky tape," "waste plastic material that happened to be on hand" or "a new kind of oil" to doctor a balky piece of equipment, but such incidents do occur. So, there's a bias against sticky tape in the original accounts, and then again in philosophical reflection. In my opinion, we have to work against this bias, and against the temptation to produce smooth symmetric theories of experimentation. Let me come back to Allan Franklin for a moment. I pointed out to him in a review (and in conversation) that the only real representation of experiment (A, as opposed to B) in his first book is the glorious photo of a mess of a laboratory on the dust jacket. The photo on the dust jacket of his second book is that of someone's laboratory notes and data. This is precisely a wrong direction, I think, in order to get a grasp on A, or real apparatus. Philosophers still need to get sticky tape on their fingers. In short, ladies and gentlemen, we need to get down and get dirty before we will have an appropriate understanding of experimentation.

References

Franklin, A. (1986), *The Neglect of Experiment*, Cambridge: Camridge University Press

________. (1990), *Experiment, right or wrong*, Cambridge: Cambridge University Press.

Pickering, A. (1984), *Constructing Quarks*, Chicago: The University of Chicago Press.

________. (1987), "Against Correspondence: A Constructionist View of Experiment and the Real," in *PSA 1986*, Volume Two, Fine, A. and Machamer, P. (eds.). East Lansing: Philosophy of Science Association. pp. 196-206.

https://doi.org/10.1086/psaprocbienmeetp.1990.2.193089 Published online by Cambridge University Press