FERMILAB-PUB-17-684

ACCEPTED

August 2017

A. FRANKLIN. *WHAT MAKES A GOOD EXPERIMENT? REASONS AND ROLES IN SCIENCE*[1]

V. Pronskikh[2]
Fermi National Accelerator Laboratory

Review of Allan Franklin, ***A. FRANKLIN. What Makes a Good Experiment? Reasons and Roles in Science. Pittsburgh: University of Pittsburgh Press, 2016. 384 pp.***

This manuscript has been authored by Fermi Research Alliance, LLC under Contract No. DE-AC02-07CH11359 with the U.S. Department of Energy, Office of Science, Office of High Energy Physics.

[1] Accepted to Annals of Science.
[2] Email: vspron@fnal.gov

Classification is an arduous and sometimes controversial undertaking because of its inherent cultural particularity. One famous literary example is the fictitious nomenclature of animals found in Jorge Luis Borges's 'The Analytical Language of John Wilkins'. According to Borges, that taxonomy was discovered by a translator, F. Kuhn, in a Chinese source and divides all animals into fourteen categories, starting with 'those that belong to the Emperor' and ending with 'those that from a long way off look like flies'. Physics of the twentieth century is a ramified full-fledged technical culture, and it should surprise no one that physics experiments can be classified in different ways, allowing certain experiments to belong to more than one category. Allan Franklin's analysis of experiments is based on his classification, which substantially grasps the essence of the experimental enterprise. Although Franklin does not offer an explicit answer to the question posed in the title, he does include a detailed examination of several important physics and biology experiments in history.

At the start of this book and along the lines of his 1981 article of the same title, Franklin distinguishes between 'conceptually good' experiments, which are important for testing or developing theories of phenomena, and 'technically good' ones, which he characterizes as 'those that measure a quantity of physical interest with greater accuracy and precision than had been done previously' (p. 2). Franklin also categorizes experiments as 'methodologically good' if they 'provided good reasons to believe their results' or as 'pedagogically important', i.e. 'those that play a didactic or explanatory role in textbooks, which they may or may not have played in the actual history' (p. 3). 'Conceptually important' experiments and 'technically good' experiments must, of course, be 'methodologically good', emphasizes Franklin. Experiments, therefore, can (and even should) belong to more than just one category. To be 'methodologically good', experiments must employ many (or all) of the ten epistemic strategies based on: various kinds of reproducibility; well-corroborated theories (both instrumental and phenomenal); statistical, manipulative and psychological techniques.

The book consists of five parts: 'Conceptually Important Experiments', 'Measuring a Quantity of Importance', 'Evidence of Entities', 'Solving a Vexing Problem' and 'Measuring Nothing'. A reader may be persuaded that most of the experiments listed are indeed of conceptual importance, even if they do not appear in the first part. Examples include the measurement of Planck's constant by Millikan, which served to test Bohr's theory (although, as Franklin explains later, was interpreted as testing Einstein's photon theory); the discovery of Higgs boson, which was evidence of the entity predicted by a dominant theory; the solar neutrino experiment, which confirmed the hypothesis of neutrino oscillations; and the Michelson–Morley experiment, which supported Einstein's relativity principle and refuted the aether hypothesis, to name a few. As one can easily imagine, if an experiment measures a quantity with a precision sufficient to test a theory and, therefore, is 'technically good', it is also 'conceptually important' indeed. To look at Franklin's unequivocally acute classification from a slightly different angle, one can say that all good experiments must be 'methodologically good' (that is, reproducible in the first place) and possess either conceptual or pedagogical (akin to classroom demonstrations of gravity phenomena) importance. Notable is how Franklin pinpoints the abundance of roles experiments play in relation to theory; it is not only the ubiquitous 'testing of an existing theory' but also investigating the subject before a theory is formulated, articulating a theory, demonstrating the existence of phenomena, presenting evidence of a new entity, and others. All these roles manifest how the gears of experiment and theory mesh, and Franklin is masterfully oriented in their diversity.

While reading the book, one can feel that the author is arguing that not all of his strategies are equally important for categorising an experiment as 'methodologically good'. An example is an admirably described disagreement between the two experiments of Thieberger and the Eöt–Wash group that involved the measurement of the hypothetical 'fifth force' — a correction to the Newtonian gravitational potential that supposedly depends on chemical compositions of the attracting bodies. Thieberger determined the correction between water and copper and claimed to have seen the effect, whereas Eöt–Wash analysed the differences between copper and beryllium and found no such ramification. Although the experimentalists relied on completely different techniques, Franklin discovered that both groups devoted significant space in their papers to discussions of their approaches to calibration, instrumentation, and elimination of various spurious effects and identified no apparent sources of errors. Moreover, the experiments have been thoroughly scrutinized by outside experts during the many years since their completion and no clear reasons for the discrepancy have been brought to light until now. However, several subsequent independent experiments conducted in the late 1980s confirmed only Eöt–Wash's result (non-existent fifth force). Finally, Thieberger himself admitted that 'some other spurious effect may have caused the motion observed', thereby renouncing the interpretation of his own observation in terms of the fifth force (p. 277).

Because the two stood in disagreement with each other, either Thieberger's experiment or that of the Eöt–Wash group had to be *ab initio* deemed methodologically flawed because they violated Franklin's first criterion, which requires the apparatus to be able to reproduce known phenomena. When several other experiments later reproduced only the Eöt–Wash team's results and not Thieberger's, the former had to be appraised as more reliable in accordance with Franklin's seventh criterion: namely, statistical arguments (understood as a more frequent occurrence of a particular outcome). Thieberger's abandonment of his effect — in fact, the explicit assertion that he had never made a discovery claim — persuaded the scientific community that an error had definitely crept into his procedure, and the supposed effect should be ignored. Quite surprisingly, Franklin argues that Thieberger's experiment should be considered 'methodologically good' (p. 300), judging it as such based on the presence of a detailed technical discussion in Thieberger's paper as well as the absence of an indication of any error revealed in the scrutiny by the community. Those arguments seem to be contestable by pointing to the argument that a scientific paper is (and this is increasingly true for contemporary physics papers) a narrative reconstruction of the experimental process that leaves many uncovered technicalities the experimentalists were unwilling or unable to unveil.[3] In addition, contemporary experiments are so technically perplexing that external experts are generally not able to scrutinize them to the last detail because of their inherent opacity.

Nevertheless, the book is unquestionably important, first of all, pedagogically, and it would be a great advantage to use it in academia to educate young prospective experimentalists and philosophers of science as to what it means to conduct a good experiment. Historians of science will definitely benefit from Franklin's diachronic

---

[3] See Allan Franklin, Shifting Standards: Experiments in Particle Physics in the Twentieth Century. Pittsburgh: University of Pittsburgh Press (2013) for examples.

approach to certain experiments that had not been regarded as important by their contemporaries but later became textbook instances of paramount ones. With clarity, elegance and decent brevity, the author presents eighteen essential experiments (only two of them in biology and the others in physics) depicting how views of their roles changed historically. Even technically overly complex experiments (such as the Higgs boson) are explained in ways comprehensible by an advanced physics student. One can concede with Franklin that 'it is fair to say then that the Thieberger and Eöt–Wash experiments are not pedagogically important' (p. 280), but one must maintain that both experiments can be valuable even pedagogically inasmuch as their discord represents a case of a controversy that accentuates the crucial yet sometimes contested role of independent reproducibility. Reproducibility appears to be an essential strategy for justification of experimental results that occupies a prominent place in Franklin's epistemology of experiment. The treatise undoubtedly deserves to become a recommended guidebook on contemporary physics and philosophy for science students to the same extent as a desk book on physics (and biology) experiments for philosophers and historians of science.