



FERMILAB-PUB-15-691

ACCEPTED

April 2015

**ALLAN FRANKLIN, *SHIFTING STANDARDS. EXPERIMENTS
IN PARTICLE PHYSICS IN THE TWENTIETH CENTURY*¹**

V. Pronskikh²
Fermi National Accelerator Laboratory

Review of Allan Franklin, *Shifting Standards. Experiments in Particle Physics in the Twentieth Century*, University of Pittsburgh Press (2013), 302 pp. \$45.98.

¹ Accepted to Philosophy of Science.

² Email: vspron@fnal.gov

Allan Franklin, *Shifting Standards. Experiments in Particle Physics in the Twentieth Century*, University of Pittsburgh Press (2013), 302 pp. \$45.98.

The book *Shifting Standards* is a valuable contribution to the literature on the history and philosophy of science and specifically to the philosophy of scientific experimentation, the discipline of which Allan Franklin is one of the outstanding scholars and founders. The central focus of the book is the contemporary shift in the norms of representation of experimental results in particle physics as well as the increasing role and drawbacks of statistical standards of acceptance for those results.

Shifting Standards begins with an explanation of the standard by which the rise of particle physics is primarily discussed—namely standard deviation, the statistical significance ascribed to a measurement. According to Franklin (2013), standard deviation is “the field’s most notable experimental standard” (ix). By virtue of the statistical nature of experimental measurements, any such measurements can be attributed to statistical uncertainty (denoted by a Greek letter σ and called sigma), which is generally as small as the number of its repetitions is large. The uncertainty (sigma) indicates how distant the measured result is from its true value and, in the case of particle physics experiments, how distant it is from the background value or the phenomena that can mimic the behavior under scrutiny. Probability theory suggests that if a measured result is just one uncertainty away from the background value (assuming the most popular, normal distribution), the probability that the measured effect is not a background event is 68 percent. However, if it is two sigmas away from the background event, the probability that the effect is genuine and not background is 95 percent. Thus, the central storyline of *Shifting Standards* is how experimentalists have changed their views of the number of sigmas required to deem valid the discovery claim of an experimental work.

Franklin's book includes a fifty-page comprehensive prologue illuminating the establishment of the "reign of sigmas" in physics, eighteen chapters describing individual experiments in physics, and a chapter titled "The Case of the Disappearing Sigmas" that scrutinizes certain cases in which experimental results claimed with high statistical significances (many sigmas) were discredited by later measurements. Franklin argues that scientists started to rely on the number of sigmas as a measure of results' credibility in the early 1960s. Before then, they usually reported values of standard deviations (sigmas) of the results in their papers but did not link their epistemic claims about the measurements with the number of sigmas. However, in the early 1970s it was remarked that "an effect of less than three standard deviations is quite insufficient" (Goudsmit quoted in Franklin, xi); Franklin cites papers published at that time claiming "observation" (which is the closest in meaning to discovery) of an effect with a four-sigma significance. "The reign of four standard deviations," writes Franklin (xxii), continued into the next decade, and in papers from the 1980s, the four-sigma results were presumed to be sufficient for an "observation" claim, although this requirement was not always explicitly mentioned. At the same time, authors sometimes regarded three-standard deviation effects as of "marginal significance." (xxiv) While not yet a formal requirement, the five-sigma criterion came into existence in particle physics in the 1990s. For example, in 1994 and 1995 the Collider Detector at Fermilab (CDF) collaboration published two papers (Abe et al. 1994 and Abe et al. 1995 cited in Franklin, xxx) on the top quark. The paper from 1994 presented a 2.8 sigma effect and was named "Evidence." The 1995 paper presented a 4.8 sigma effect and was called "Observation." Nevertheless, even at that time the five-sigma standard "was certainly not agreed upon or enforced" (xxxii). The five-sigma standard was issued as an instruction by the editors of major physics journals around 2003 and enjoined by many collaborators for submission of "Observation"

papers (many papers presenting four or fewer sigma measurements were named “Evidence”). The same five-sigma rule had already become a well-established standard when the first Higgs boson discovery papers were published in 2012 (Chatrchyan et al. 2012 cited in Franklin, li). Franklin poses a question of whether the discovery claims of the past experiments relying on fewer than five sigmas might be reconsidered retroactively.

A case study of the “disappearing sigmas” (Franklin 2013, 212) takes a central place in the book. In 2003, a number of experimental groups found a “pentaquark,” an elementary particle composed of five quarks with significances nearing five sigmas. These results were supported by certain theoretical predictions. If combined, their aggregate statistical significance totaled seventeen to twenty sigmas. (220) In order to obtain those numbers, first the probabilities published by the nine different positive “pentaquark” experiments (each claiming a four or five-sigma significance) were converted to probabilities that their observations were statistical fluctuations; the uncertainties were included in the probabilities. Those probabilities were then multiplied in accordance with statistical rules to yield the resulting aggregate probability, which was subsequently transformed into the number of sigmas. (Franklin 2015) Nevertheless, in 2006 three other experimental papers were published refuting the existence of “pentaquarks.” Eventually, the community concluded that the observation claim of “pentaquark” was “a false alarm.” (Wohl quoted in Franklin 2013, 219) The book contains a reference to another twenty-sigma claim of a neutral boson discovery that was not confirmed by later experiments, that of (Maglich 1971, cited in Franklin 2013). (xvii) This case study raises a number of questions, and Franklin searches for the answers to them throughout the book.

First, the sigma criterion is usually applied to only selected amounts of data (cuts) chosen by data analysts on grounds known only to them. However, if the selections were different, the

number of sigmas would be essentially different as well (the so-called Look Elsewhere Effect). Second, in addition to statistical uncertainties, systematic errors must be accounted for (and the undermentioned bias discussion seems to be very relevant here). Systematic errors can include either unknown background phenomena or unknown regimes of functioning apparatuses, and all of them can be regarded as ignorance errors. Franklin emphasizes that knowledge of the details of the experimental and data analysis procedures are essential for understanding how reliable results are and what claims can be made on their basis; they cannot be reduced to solely statistical arguments. He proposes a numbers of epistemic strategies that have to be involved (and, therefore, reflected in writing by the experimentalists) in addition to statistical arguments, namely, reproducing known and new effects, checking the internal consistency of results, using well-corroborated theories for both the design of apparatuses and interpreting the results.

In this connection, Franklin devotes a chapter to a discussion of good and bad data as well as exclusion and selectivity procedures. What he calls good data is data not only produced by a properly working apparatus but also purified from the background effects, discussed above, that are capable of mimicking the genuine one. Generally, to eliminate background effects, the data acquired in an experiment undergo an analysis that employs certain theoretical frameworks explicitly. However, Franklin's selectivity implies singling out such good data even before starting such analysis by designing or tuning apparatuses. He writes, "Experiments rarely work properly when initially turned on" (127). On the other hand, he defines bad data as that resulting from a failing apparatus or misrepresentation. Exclusion means elimination of such data. Taking as examples the nuclear spin and Millikan's experiments, Franklin remarks that in the course of experiment, experimentalists usually find or envision flaws and make changes to apparatuses to eliminate them. However, that practice is rarely reflected in published papers. The latter finding is

consistent with Peter Galison's observation that "reading an article, one could conclude that an effect would follow from an experimental setup with the inexorability of logical implication," although only the experimentalist knows the actual strengths and weaknesses of his or her "machines, materials, collaborators, interpretations, and judgments" (Galison 1987, 244).

Franklin focuses on issues such as experimenters' biases in cases where the results agreed with high-level theory, experimentalists' presuppositions, or previous measurements, thus reiterating the importance of non-statistical errors. He provides detailed information on each experiment's apparatuses and data analysis procedures based on the original publications in major scientific journals. This approach can be criticized based on the aforementioned claim that experimental papers rationally reconstruct the actual procedures, leaving many essential aspects undeclared. Nevertheless, Franklin's claim seems sound enough. First, due to Franklin's wealth of experience in experimental physics, it is difficult to doubt his ability to read between the lines of the papers under scrutiny. Second, this finding is usually relevant for contemporary big science papers whereas the book comprises experiments dating back to Millikan's paper of 1911, which has not been notably criticized for constructing the logic of narrative. For example, in early papers authors often assessed personalities of previous authors and made critical judgments about their works based on their perceived personality flaws (Franklin 2013, 8).

Franklin points out that the meaning of being an author of an experimental paper has also changed as has the number of the authors of a paper (from one in the 1911 Millikan experiment to two thousand in the Compact Muon Solenoid (CMS) experiment at CERN). He unveils a revealing story from the end of the 1950s of a young scientist who had substantially contributed at many stages of an experiment but was not accorded the status of author by the senior authors of the paper because he "did not have sufficient knowledge to give a talk about the experiment" (4). According

to Franklin, in contemporary collaborations only five to twenty members have access to the data analysis (leading to experimental results) and, therefore, know the analysis sufficiently to give talks on the experiment, while contributors to design and running shifts, physicists, and engineers, are on the authors' list, too. The speechless others in the collaboration are engaged in the construction of apparatuses, computer programs, and other technical activities but do not have permission to analyze data or present the collaboration in general talks at outside institutions and conferences (though they can be allowed to report their particular technical achievements). Thus, Franklin highlights tremendous changes in authorship policies between the middle of the twentieth century and the twenty-first century due to the increase in authors' lists and their epistemic stratification. This controversy is not an extensive part of the book but is worthwhile.

Shifting Standards is a solid work on experiments in physics and experimental epistemic strategies, revealing transformations of epistemic norms in particle physics and portraying them as seen through a physicist's lens. It is abundant in technical details but not so overwhelmingly that the subject matter escapes an observant reader. The book can be recommended not only to historians and philosophers of science but also to reflexive practicing scientists as a catalyst to help them examine themselves.

References

Galison, Peter. 1987. *How Experiments End*. Chicago: University of Chicago Press.

Franklin, Allan. 2015. Private communication.