

Franklin's Field Guide to Scientific Experiments

Nora Mills Boyd*†

Allan Franklin, *What Makes a Good Experiment? Reasons and Roles in Science*. Pittsburgh: University of Pittsburgh Press (2016), 384 pp., \$55.00.

Some manuscripts affect our field like the arrival of an alien spaceship (Kuhn's *Structure*) or sound out like a call to join the revolution (Feyerabend's *Against Method*). Allan Franklin's 2016 book exemplifies a different style of philosophy entirely. The contribution of this work is that of tending, stoking, nurturing—of keeping alive a tradition and of continuing the process of thinking through an important matter. What makes a good experiment? Franklin concludes: "I do not have an answer to the question" (296). He names Mendel's experiments in plant hybridization as "the best experiments ever done" but suggests that "there is no simple algorithm for evaluating or ranking good experiments" (304, 306). At the end of the day, Franklin claims to leave readers "to make their own judgements" (306).

In fact, he provides somewhat more in the way of an answer to the work's guiding question. For Franklin, "methodological goodness" is a necessary condition for an experiment to be good, and adding "to scientific knowledge" or being "helpful in acquiring that knowledge" are further desiderata (297, 300–301). The aim of the book is to present the details of actual experiments so that the reader can see concretely in what such goodness consists.

Reading Franklin's book is not unlike being taught by a connoisseur to use a field guide, such as *The Sibley Guide to Birds*, to identify notable avian characteristics. "Attend to the silhouette, the flight pattern, and the field marks on the head and on the wing" an experienced birder might advise.

Received March 2017.

*To contact the author, please write to: Department of History and Philosophy of Science, University of Pittsburgh, Pittsburgh, PA 15260; e-mail: nmb58@pitt.edu.

†The author is grateful for helpful feedback from John Norton, Aaron Novick, Kathleen A. Creel, Siska De Baerdemaeker, David Colaço, and Allan Franklin on this review.

Philosophy of Science, 84 (July 2017) pp. 586–594. 0031-8248/2017/8403-0010\$10.00
Copyright 2017 by the Philosophy of Science Association. All rights reserved.

Likewise, Franklin encourages his apprentice readership to attend to the characteristic roles that experiments play and the reasons that scientists rely on in arguing for the credibility of their findings. Franklin gets us started using his field guide by working through case studies for us—carefully demonstrating the identification of different characteristics of experiments (“Look *there* at the crown stripe and whisker mark on that white-throated sparrow!”). After several examples, evidently satisfied that we ought to have gotten the hang of it by now, Franklin sets his readers loose with the guide, to try our own hands at recognizing good experiments from their field mark equivalents.

This book is obviously relevant for any philosopher of science working in the epistemology of experiment, but I would also recommend it to anyone who needs their view of experiments as testers of theory updated with a more nuanced and variegated picture. Below I outline the contents of the field guide portion of Franklin’s work (substantive engagement with the case studies themselves is beyond the present scope). I conclude by articulating what I think is an important sense in which Franklin’s book is a valuable philosophical contribution and pointing out one place in which I think Franklin failed to fully capitalize on this contribution.

In the introduction, Franklin offers three classificatory frameworks that he deploys throughout the book: *roles* of experiments in science, ways in which an experiment can be *good*, and *strategies* used in arguing for the correctness of an experimental result. According to Franklin, a single experiment can play more than one role, be good in more than one way, and make use of more than one strategy in supporting the credibility of its result. I will discuss each of these frameworks briefly in turn.

Roles. The collection of roles builds on those that Franklin identified in his 1981 paper titled “What Makes a ‘Good’ Experiment?” Ecumenical as always, Franklin demurs at the possibility of providing a complete list of the roles of experiment in science once and for all, claiming that he does not “believe that this list of the varying roles that experiments play is exclusive or exhaustive” (2). Although Franklin introduces the various roles (I count 12) by peppering them naturally in his prose, I submit that for the most part they can be usefully grouped into three rough categories of experiment types. These types are what I will call “theory modifiers,” “theory fodder,” and “methodological advance.” Roles that belong to the theory modifiers category include, for instance, *deciding between competing theories* and *helping to articulate an existing theory*. In contrast, experiments that serve as theory fodder include exploratory experiments used in *investigating a subject for which a theory does not exist* and those used for *measuring quantities of physical interest*. In the 1981 paper, the roles Franklin discussed fell within these first two categories. He argued that “a ‘good’ experiment is one which bears a

conceptually important relation to existing theories . . . or calls for a new theory. It must also measure the quantity of interest to sufficient accuracy and precision” (Franklin 1981, 372).

Roles in what I am calling the “methodological advance” category did not appear in the 1981 paper. These are roles that experiments can play largely independently of their relation to theory testing or development. Franklin names at least two in the work under review: “give an incorrect result but demonstrates that the quantity of interest can be measured” (enabling experiment) and “demonstrate a successful new experimental technique” (2). He identifies early experiments testing atomic parity violation as an example of “enabling experiment” (2 n. 1). The construction of a neutrino beam serves as an example of demonstrating a successful new experimental technique (2).

A final role for experiments that Franklin notes is to “have a life of their own, independent of high-level theory” (2). This role remains without explanation or exemplar in the 2016 book. Franklin and Perović do name examples of this role explicitly in the “Experiment in Physics” entry of the *Stanford Encyclopedia of Philosophy* (*SEP*; 2016, sec. 2.1), including several from Hacking (1983): “Carolyn Herschel’s discovery of comets, William Herschel’s work on ‘radiant heat,’ and Davy’s observation of the gas emitted by algae,” work “on Iceland Spar by Bartholin, on diffraction by Hooke and Grimaldi, and on the dispersion of light by Newton,” adding to these “the nineteenth century measurements of atomic spectra and the work on the masses and properties on elementary particles during the 1960s” (see also Franklin 1993, 114–15). However, the relationship between the role articulated by this list of experiments in the *SEP* article and the *having a life of their own* role in the 2016 book is not completely clear. At the end of the *SEP* section, Franklin and Perović (2016) state: “In all of these cases we may say that these were observations waiting for, or perhaps even calling for, a theory,” yet in the book these two roles (“call for a new theory” and “have a life of their own, independent of high-level theory”) are separated out (1–2).

Without further details it is difficult to tell whether *having a life of their own* belongs in one of the categories I have already specified or perhaps requires a category of its own. In my estimation, Bartholin’s work on Iceland spar served as theory fodder by *investigating a subject for which a theory does not exist*. In contrast, measurements of masses and properties of particles in the 1960s could plausibly be classified either as theory modifiers (e.g., *helping to articulate an existing theory*) or theory fodder (e.g., *measuring quantities of physical interest*) depending on the particular measurements considered. Yet there is reason to think that Franklin has something else in mind by *having a life of their own*—something like experimental inertia. Franklin (1993) describes the continuation of fifth force style gravita-

tional research in these terms: “The Fifth Force may indeed be dead, but work continues. Experiments do seem to generate a life of their own” (105).

In this case Franklin explains the continued experimentation by appealing to such factors as the familiarity experimenters gained with their apparatuses and cost effectiveness (1993, 123). This suggests a role that I do think would be importantly distinct from the others Franklin outlines: a kind of self-perpetuating experimental culture, motivated primarily by allegiance to certain instruments and techniques rather than interaction with high-level theory. If this is the gist of Franklin’s suggestion then it would be interesting to scrutinize the implicated case studies further in order to evaluate whether such a description is indeed apt. Experimental activity in this sense would resemble (or perhaps become) artisanal or technical activity. I would worry, however, that attempting to provide an epistemology of such activity would be a project significantly different from the rest of Franklin’s corpus.

Good and Methodologically Good. The second classificatory scheme that Franklin uses concerns the ways in which an experiment can be good. Among these, according to Franklin, are “conceptually important,” “technically good,” “methodologically good,” and “pedagogically important” (2–3). Again, Franklin warns, “I do not believe that either of these lists exhausts the roles that experiment plays in science or the ways in which an experiment can be good” (3).

The third classificatory framework is Franklin’s familiar epistemology of experiment (also enshrined in the *SEP* entry)—“strategies that can be and are used to argue for the correctness of an experimental result” (3). In other words, Franklin’s epistemology of experiment serves to explicate the ways in which an experiment can be methodologically good, that is, provide “good reasons for belief in their results” (3). Again, Franklin claims that the list of strategies that comprise his epistemology of experiment is “neither exclusive nor exhaustive” but that “the use of such strategies is . . . necessary to establish the credibility of a result” (4).

I will not reproduce the list of 10 strategies that currently comprises Franklin’s epistemology of experiment here (e.g., 4). Suffice it to say that I believe these can be grouped into more coarse-grained bins as well, in this case according to the rough phase of experimental process to which they correspond: “preparation” (e.g., calibration), “analysis” (e.g., statistical arguments), and “cross-checking” (e.g., eliminating sources of error). I leave it as an exercise for the reader to augment Franklin’s list with further strategies, to take aim at those that he has already identified, or to illuminate further structure in this taxonomy.

The Experiments. *What Makes a Good Experiment?* is divided into five parts, corresponding to different roles: significantly changing theory, mea-

asuring an important quantity, providing evidence for entities, solving a vexing problem, and null experiments. Franklin does not explicitly discuss why these select roles are treated in detail, given the relative abundance of roles named in the introduction, or how “solving a vexing problem” relates to those roles. The chapters belonging to each part discuss an experiment or set of closely related experiments exhibiting the associated role. Out of 18 chapters on experiments, two are dedicated to experiments in biology (Mendel and Meselson-Stahl), and the rest to physics—mostly nuclear and particle physics. In each case, Franklin walks the reader through the details of the experiment or experiments in question and then provides a discussion section in which he tacks elements of his classificatory schemes onto the case at hand. (I cannot help but take this opportunity to register a complaint, which is not so much directed at Franklin as at the University of Pittsburgh Press. The book is laced with an abundance of endnotes, to the sustained consternation of the committed reader. The trouble is that the contents of the notes are unpredictable. Some are throwaways, some lovely jokes, some references, some fill in further details or historical context, and some are important philosophical contributions. Using footnotes instead would allow readers to satisfy their curiosity with minimal interruption of smooth reading.)

With the standout exception of chapter 4 in which Franklin enlists the initial proposal for the experiment and the laboratory notebook (57 n. 1), details regarding the experiments are almost always furnished by references to a single published scientific paper as the primary source. Although a long-standing practice for Franklin, this aspect of his approach makes me nervous. To do the history of an experiment—to substantiate a reconstruction of the methodologies actually employed in practice—would require investigating well beyond the facades of scientific publications into the murky subterranean realm of laboratory records, correspondence between scientists, interviews, and iterations of paper drafts. What we get in published scientific articles are the carefully crafted presentations of experiments structured and emphasized according to how the scientists see fit. The subtle pathways actually tread in the laboratory to arrive at a result can be hidden from view or dramatically recast in the final, public analysis. Franklin’s position is that in cases in which he has explored historical material beyond publications, such as correspondence between scientists, he has not found anything that would cause him to change his opinion of the experiment in question, stating that “for an epistemologist the public science is, I believe, far more important” (e-mail message to author, December 18, 2016). However, I think this is a dangerous induction. Perhaps one need look no further than the “replication crisis” in psychology to see why access to material beyond final publications could be epistemically relevant.

So it cannot be that what Franklin has given us in *What Makes a Good Experiment?* is a collection of historical case studies of scientific method-

ology in situ. Instead, I suggest that this book is best construed as providing patient discussion of the sort of reasons that scientists present to one another in the daylight of peer-reviewed publications in the service of a variety of epistemic aims. Thus, it is more like using a field guide indoors in a natural history museum than out in wild meadows. This is certainly in itself a worthwhile project because peer-reviewed publications are the primary medium through which scientific results are absorbed across communities of researchers. It is rare indeed for groups of scientists to have access to the internal documents and practices of other groups. However, the limitations of Franklin's approach in this book point the way toward a complementary subject that would also be worth investigating with similar care: the reasoning that scientists employ on their way to a publishable result.

A Monster. What is the philosophical value of having a field guide to scientific experiments? Perhaps the most obvious value is that of rendering the complex landscape of scientific practice more intelligible. With roles, ways of being good, and strategies made explicit, we can sharpen both our perception of science and our tools for improving it. Making the epistemology of experimentation explicit also allows us to see more clearly when things have gone wrong, when the reasoning of scientists strays from what is epistemologically responsible.

Just such an example leaps out at the reader in Franklin's book, specifically in chapter 17, "A Tale of Two Experiments: Is There a Fifth Force?" The tale involves Peter Thieberger's differential accelerometer experiment, results from which supported the existence of a fifth force, and the Eöt-Wash group's torsion pendulum experiment, results from which ruled out the existence of such a force (see also Franklin 1993; Franklin and Fischbach 2016). What should we, and the physics community, make of the fact that two experiments produced *prima facie* discordant results? This is precisely the sort of case that cries out for the resources of an explicit epistemology of experiment of the sort that Franklin champions.

What should certainly not be acceptable from the viewpoint of the epistemology of experiment is if this sort of evidential discord were decisively resolved by forces other than reasoned argument. Yet this seems to be precisely what has occurred in this case. According to Franklin, Thieberger did everything right—his experiment was methodologically good—and yet "a decision was made that Thieberger's result was wrong and that the Eöt-Wash result was correct" despite the fact that "after several years of scrutiny, and even to this day, no one has found an error in either [the] experiment or in its analysis" (277). On what basis was this decision made? According to Franklin, "Thieberger and the rest of the physics community were persuaded by an overwhelming preponderance of evidence that there was no fifth force" (277). Franklin recently reflected: "As an experimentalist you don't make

a lot of money looking for errors in other scientist's work" (e-mail message to author, December 18, 2016). If we read Franklin's book as a field guide, then Thieberger's experiment appears as a monster that has somehow crept into his collection. As he presents the experiment in chapter 17, Franklin seems both to acknowledge that the appraisal of Thieberger's experimental result falls outside of Franklin's own epistemology of experiment and to tacitly condone that appraisal. This is surprising. Borges-like, Franklin invites us to believe in the monster by failing to protest stringently against the physics community's apparently ungrounded pronouncement on the result. What has happened here? How did methodological goodness and the credibility of a result become decoupled in this case? And why ain't anyone reaching for their pitchfork?

One way this tension could be resolved would be to add a strategy to Franklin's list that would reflect the reasoning employed in this case. For instance, perhaps we want to say that scientists can argue for the credibility of a result by *amassing a preponderance of evidence against results that are discordant with it*. Does this strategy rightly belong in an epistemology of experiment? Franklin (1993) says as much: "It is not necessary to know the exact source of an error in order to discount or to distrust a particular experimental result. [Its] disagreement with numerous other results can, I believe, be sufficient" (109 n. 122). While I am sympathetic to something in this neighborhood, I think we need to proceed with caution. It strikes me that simply amassing a preponderance of evidence in one camp does not by itself furnish good reason to discount discordant results and thereby increase the credibility of the mass of evidence. The minority view could be the right one. So when is it reasonable to discount a discordant result?

If Thieberger's result is indeed wrong, then it should be the case that something was overlooked in the appraisal of Thieberger's experiment that, if uncovered, would explain the discrepancy. Of course it could be that Thieberger's experiment was methodologically good, in the sense that no one could have reasonably expected him to do anything more or better to justify his result, and nevertheless in fact be the case that some mistake or systematic error is responsible for the discordant evidence (cf. 278). Moreover, it could very well be that the overlooked mistake or unaccounted for systematic error will unfortunately never be revealed. In principle, every detail of Thieberger's experiment is potentially implicated, but sufficient records may not survive to trace our way through them all. The problem element may have already been lost to history. What we may reasonably say about Thieberger's result depends importantly on the accessibility of information about the production of that result. Without sufficient information to tell whether some mistake or systematic error could be responsible for it, we (and the physics community) ought to be agnostic about Thieberger's result, rather than allowing the result to be pronounced "wrong" without justification. In-

terestingly, according to Franklin, Thieberger himself “thought he had made an error, even if he didn’t know what it was” (e-mail message to author, December 18, 2016). Yet while it seems perfectly reasonable, given an overwhelming preponderance of evidence produced by experiments that are themselves methodologically good, for the physics community to proceed on the (fallible) assumption that there is no fifth force, this does not license white-washing the result of the discrepant experiment as definitely wrong (for unknown reasons).

This case makes it clear that we need an epistemology of experiment that can deal with anomalies and discordant results responsibly, that is, without tacitly condoning dogmatism. There are many examples apart from Thieberger’s of experiments that have disrupted the harmony of bodies of evidence. As Franklin himself puts it: “it is a fact of life in empirical science that experiments often give discordant results” (2002, 35). One of my personal favorites is the annual modulation signal detected by DAMA and its successor experiment DAMA/LIBRA, which the experimenters have interpreted as a detection of galactic dark matter since the late 1990s (cf. Bernabei et al. 2013). This result has not been widely accepted by the physics community, but, to my knowledge, the cause of the signal remains mysterious. Every effort should be made to disenchant such beasts.

In the case of the DAMA/LIBRA dark matter signal—as in the case of Thieberger’s fifth force—it would be epistemically problematic to reject the result without accounting for its origin. To do empirical science is to commit to constraining one’s understanding of the world with the available evidence. To pass over discrepant results would be to fundamentally undercut this project. We are burdened with discrepant signals, anomalies, and Sasquatch sightings until they have been reinterpreted in an epistemically responsible way. To sit easy with the fact that “a decision was made” in absence of such justification would be to concede the philosophical stance to the sociological. Despite his recent truce with Harry Collins (Franklin and Collins 2016), I suspect that Franklin, pace chapter 17, would still find such a concession deeply disagreeable.

REFERENCES

- Bernabei, Rita, et al. 2013. “Final Model Independent Result of DAMA/LIBRA-phase1.” *European Physics Journal C* 73 (12): 2648.
- Franklin, Allan D. 1981. “What Makes a ‘Good’ Experiment?” *British Journal for the Philosophy of Science* 32 (4): 367–74.
- . 1993. *The Rise and Fall of the Fifth Force: Discovery, Pursuit, and Justification in Modern Physics*. New York: American Institute of Physics.
- . 2002. *Selectivity and Discord*. Pittsburgh: University of Pittsburgh Press.
- Franklin, Allan D., and Harry Collins. 2016. “Two Kinds of Case Study and a New Agreement.” In *The Philosophy of Historical Case Studies*, ed. Tilman Sauer and Raphael Scholl, 95–121. Boston Studies in the Philosophy of Science 319. Dordrecht: Springer.

- Franklin, Allan D., and Ephraim Fischbach. 2016. *The Rise and Fall of the Fifth Force: Discovery, Pursuit, and Justification in Modern Physics*. 2nd ed. Dordrecht: Springer.
- Franklin, Allan D., and Slobodan Perović. 2016. "Experiment in Physics." In *Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta. Stanford, CA: Stanford University. <https://plato.stanford.edu/archives/win2016/entries/physics-experiment/>.
- Hacking, Ian. 1983. *Representing and Intervening*. Cambridge: Cambridge University Press.