Introduction

> I prefer you to practise by drawing things large. . . . In small drawings every large weakness is easily hidden; in the large the smallest weakness is easily seen.
>
> Leon Battista Alberti ([1435] 1966, 94)

Big science presents a big opportunity for methodologists. With their constant meetings and exchanges of e-mail, collaboration scientists routinely put their reasoning on public display, long before they write up their results for publication in a journal. For philosophers who wish to understand scientific evidence and scientific reasoning, such thinking out loud can be a rich source of information. (And given the expense of mounting an experiment in a field such as high energy physics, why not get as much as possible out of it? For all the interesting physics that emerges, the episode may also yield results for philosophical, social, and historical studies. Think of it as scholarly recycling.)

So here is a study of big science: the search for the elementary particle known as the *top quark*, the last of the six quarks of modern physics' "standard model" to be experimentally confirmed, by the Collider Detector at Fermilab (CDF) collaboration. CDF is certainly *big*. By the time that they announced their "observation of the top quark," CDF had swelled to a size of approximately 450 physicists. But in virtue of what does CDF's enterprise constitute *science*?

Without here attempting to specify necessary and sufficient conditions for science, I do wish to suggest that a central aim of a scientific enterprise is to determine what general claims about the world can be supported with evidence. That is, scientific enterprises are evidence-oriented.

At first glance, such a claim may seem overly simplistic, suggesting that scientists engage exclusively in testing theories and evaluating the outcomes of such tests. The chapters to come should dispel any such notion. Theorizing on a speculative level, scheduling meetings, designing experimental apparatuses, raising funds, performing feasibility studies, developing review and oversight procedures, arguing over the choice of persons to carry out such procedures, running computer-aided simulation of data, calibrating instruments, and gossiping about one's competitors – these are all part of the story that I am about to tell. Yet throughout this story, we will find that, even when it does not take center stage, evidence plays an important role in the unfolding of events. The exact nature of that role depends on the context. Questions about evidence, or the potential for evidence, enter in one way during the development of speculative physical theories (Chapter 1) but in quite a different way during the design of a new detector (Chapter 2) or when deciding on the best means for analyzing data (Chapter 3). Evidential considerations move into the spotlight when, with the theory, instrument, and analysis in place, scientists ask themselves whether they do indeed have evidence for a specific claim, or, to turn the coin to its opposite side, for what specific hypothesis they can say that they have evidence (Chapters 4 and 5).

I chose to study this particular episode because of the detailed publication on which it centers, the "Evidence for Top Quark Production" paper published in 1994 by the CDF collaboration, and because of the interesting methodological issues raised by the debate within CDF over their evidence claim. My choice is problematic on two counts.

First, by focusing on CDF's history and search for the top quark, I slight the D-zero collaboration, which joined CDF in announcing in 1995 the "Observation of the Top Quark." D-zero is also situated at Fermilab, and although they are a younger collaboration than CDF, they are just as large and could easily serve as the subject of a lengthy study such as this. However, neither of the 1995 "Observation of the Top Quark" papers involved the kind of methodologically significant controversies generated by CDF's "Evidence" paper. I wish to emphasize, however, that *this is not the history of the discovery of the top quark*. No work that gives such slight attention to the D-zero collaboration could claim that title. My decision to emphasize CDF's efforts should not be read as any kind of judgment regarding the relative worth or importance of the work done by the two collaborations.

Second, this episode is not in any way representative of a class of things such as "scientific experiments," or even "particle physics experiments." In many ways, the case I have chosen is extreme: extreme in its complexity, scale, and level of contentiousness as an experimental result. I do not propose here to reach philosophical conclusions by generalization from this single episode. My approach is much more piecemeal.

Instead, in the first five chapters, which are primarily historical in their emphasis, I seek to discover what *can* happen in the natural sciences, rather than what typically does happen, by examining some aspects of physicists' work on the top quark. Thus, this study serves as a source of suggestions for further historical study. Also, these historical chapters narrate the progress of a number of controversies that arose during CDF's search for the top quark and situate those controversies within the specific context of CDF. The fact that some methodological issues in experimental science are specific to particular contexts is relevant to the ensuing philosophical discussion.

That philosophical discussion occupies the last two chapters, in which I adopt a certain theoretical framework regarding evidence, the "error statistical" theory (Mayo 1996). According to the error statistical theory, an experimental test yields evidence for a hypothesis when (or insofar as) the result of the experiment not only fits the hypothesis but is also of a kind that would be very improbable if the hypothesis in question were not true. The latter requirement is the *severity* requirement. A convincing argument for a hypothesis must therefore convince the audience that the hypothesis has passed a severe test. In my concluding chapters, I seek to explicate the *concept* of evidence at the center of the error statistical theory, particularly with respect to two issues: (1) In what sense, if any, is that concept objective? (2) Does the error statistical concept of evidence relate in any useful way to belief? In addition, I pose two main questions for the error statistical *theory* of evidence: (1) Does it provide the resources for explaining the relevance of those considerations on which debate over CDF's claim to have evidence for the top quark turned? (2) Does it provide the resources for proposing potentially useful strategies for avoiding methodological difficulties of the sort encountered by CDF, and for coping with them once they have arisen? The ability to accomplish these two aims constitutes, I propose, a necessary condition of adequacy for any theory of evidence. My hope is that the historical details presented in the earlier chapters will provide the reader with sufficient information to evaluate my own answers to these questions.

Chapter 1 concerns the introduction into the Standard Model of the fifth ("bottom") and sixth ("top") quarks by Makoto Kobayashi and Toshihide Maskawa in 1973 (Kobayashi and Maskawa 1973). In particular, I discuss the largely forgotten background to Kobayashi and Maskawa's widely acknowledged work. Their work had its roots in the school of theoretical physics at Nagoya University over which Shoichi Sakata presided. Sakata and several other Nagoya physicists were committed to dialectical materialism, which served as the basis for a methodology for advancing physical theory. Sakata and colleagues invoked that methodology in turn in proposing a theory of the structure of matter – the "Nagoya Model" – that anticipated more famous later developments. The Nagoya model then came to serve as the framework in which a number of Japanese physicists interpreted the results of cosmic ray experiments suggesting the existence of a new fundamental particle, and it was this apparent extension of the class of elementary particles that prompted Kobayashi and Maskawa toward the theoretical studies that led to their introduction of an entirely new "generation" of matter, including what would later come to be called the top quark. In restoring part of the history of the developments that led up to the search for the top quark, I also discuss the role of philosophical commitments in physical theorizing.

Chapter 2 chronicles the origins of CDF and the design and construction of their remarkably complex detector. The story told here details how the group that became CDF chose a detector design from among many different possibilities, against a background of constraints that were evolving even as the members of the collaboration were attempting to finalize the blueprint for their detector, and while they attempted to recruit both talented physicists and funding to support a rapidly growing enterprise. I argue that during the ten years spent planning and building the detector before gathering any data at all, CDF's aims were quite general. They created a resource for an extended experimental program, capable of supporting tests of a number of general types of physical hypotheses, without locking themselves into the pursuit of specified theoretical claims. They expected the detector to accommodate a variety of experimental pursuits, in virtue of the detector's ability to measure certain quantities, the interest of which was largely independent of the particular theories to be explored. Allan Franklin has described how experimenters exhibit "instrumental loyalty," shaping their experimental programs to utilize already operable, well-understood instruments (Franklin 1997). The early years of CDF show the other face of instrumental loyalty, where experimentalists set out to build a device capable of rewarding such loyalty, even though the precise course of experimentation remains to be determined.

Crucial to CDF's top quark evidence claim was an upgrade of the detector resulting in better discrimination between particle decay events involving top quarks and other particle events yielding similar results. The improvement required CDF to insert into the heart of their detector a new component based on technology never before tried in such a setting. I explain in Chapter 3 how a group of Italian physicists interested in applying this new technology won over their colleagues by using persistence and ad hoc arguments and by deferring to physicists at another institution to solve a recalcitrant technical problem. Also helpful was the fact that CDF's "detector philosophy" was compatible with the proposed detector component. Peter Galison has discussed how such detector philosophies shape experimental programs (Galison 1997b), and I contrast my interpretation of that phenomenon with his. This episode shows how the development of new techniques of detection and instrumentation technologies exhibits a certain degree of autonomy from specific experimental and theoretical aims.

Also important to Chapter 3 is the struggle within CDF as physicists developed strategies for finding the top quark. Political intrigues and social forces figured prominently, but not simply by making some approaches look convincing and others not convincing, as a strictly "social constructivist" reading might suggest (see, e.g., Pickering 1984). Rather, alienation from colleagues made it difficult for some working on complicated methods to get others to study their proposals. Their analyses may or may not have worked, but it is possible that these methods were not adopted because not enough of the collaboration devoted the effort needed to evaluate them. (Others in the collaboration insist that, beyond the "political" difficulties of some of these proposals, the proposed methods simply did not work in a reliable way, even when one did take the time to study them carefully.) Although one cannot deny that the internal social dynamics of the CDF collaboration influenced the careers of these proposals, such social forces served not to *determine* or *constitute* the validity or invalidity of the techniques proposed, but rather to encourage or set obstacles to the study needed for assessing their validity.

Based on data collected during a period known as "run Ia," CDF submitted to the journal *Physical Review D* a lengthy account of the analysis that led to their evidence claim (the "Evidence" paper). I describe in Chapter 4 the months that CDF spent writing that article, and the many disagreements and controversies that boiled within the collaboration. Many CDF members criticized the interpretation of the data favored by group members directing the writing of the Evidence paper. I argue that part of the function of the published research report is to facilitate just such criticism of one's own work by those best situated to discover certain types of errors. If evidence is produced by exposing hypotheses to severe tests during an experiment, then writing a paper after the experiment can become the pursuit of experiment by other means. While writing, experimenters can test the various assumptions on which they rely in arriving at an evidence claim, often by reanalyzing existing data in different ways. Such a conception of the aim of writing a research report accounts, in part, for the fact, decried by some, that scientific papers typically provide misleading, cleansed narratives of the research process (Medawar 1964). My argument also adds a new dimension to Thomas Nickles's observation that in writing a research report "scientists are not writing *about* science; they are *doing* it" (Nickles 1992, 96).

I turn to the aftermath of the Evidence paper in Chapter 5. After a brief shutdown, CDF resumed taking data during "run Ib," and after another year, both CDF and D-zero announced that they had "observed" the top quark. In the meantime, some in CDF used new data and new methods of analysis developed *after* the Evidence paper to cast doubts on some aspects of the earlier results. Some CDF members believed that the earlier tests were biased, resulting in a misleading assessment of the statistical significance of their data. Subsequent developments, they said, yielded two sources of support for their suspicions: (1) reanalysis of existing run Ia data using newly developed algorithms, and (2) application of similar algorithms to larger bodies of data from both run Ia and run Ib. I argue that (1) constitutes an example of a "robustness analysis" (Campbell and Fiske 1959; Wimsatt 1981; Culp 1995) used to test the reliability of a result, while (2) reflects an awareness that, as data accumulate, genuine effects should yield more impressive statistical significance calculations. Both kinds of testing show that evaluating evidence may require further empirical (in some cases diachronic) investigations, as one would expect if evidence claims are indeed empirical (Achinstein 1995; 2001) and subject to additional testing.

With the historical narrative complete, Chapters 6 and 7 are more strictly philosophical. In Chapter 6 I discuss the error statistical theory in more detail. First, I develop an outline of an error statistical model of CDF's experiment as reported in the Evidence paper, utilizing a hierarchy of models connecting data to high-level theoretical claims via a (partial) model of the experiment itself. I further argue that the error statistical theory supports "local" comparative assessments of evidential strength (comparisons of different outcomes of a given experiment as evidence for a given hypothesis, or comparisons of a given experimental outcome as evidence for different members of a family of hypotheses), and that it also supports a classificatory concept of evidence, enabling one to make judgments that certain results do or do not constitute evidence for a particular hypothesis. However, the error statistical theory does not enable the kind of "global" comparisons (comparing the strength of e as evidence for h, and e' as evidence for h', for any e, e' and h, h' whatsoever) that would be required for a quantitative measure of evidential strength. Nevertheless, both the comparative and classificatory concepts of evidence described by the error statistical theory are fully objective in the sense that whether an experimental outcome is error statistical evidence for a hypothesis, or stronger evidence for one hypothesis than for another, is independent of whether any person believes it to be such. Furthermore, if a result is error statistical evidence for a hypothesis in the classificatory sense, then it constitutes a reason to believe that hypothesis independently of whether any person believes it to be such, and independently (in an epistemic sense) of the epistemic state of the experimenter or anyone else (see Achinstein 2001).

Finally, Chapter 7 addresses problems of bias in experimental testing and revisits the question of objectivity. A recurring point of controversy within CDF was the problem of biasing a test toward a favored outcome by changing one's data selection criteria in the face of knowledge of the data to which they will be applied, a phenomenon known to particle physicists as "tuning on the signal." Precisely in order to avoid such problems, experimenters in many contexts insist on "predesignating" the features of their testing procedure prior to acquiring any knowledge of any data. I compare the predesignation requirement to the requirement of "novel predictions" that has long preoccupied philosophers of science. I argue that if we take the experimenter's concern to be, not providing novelty as such, but ensuring that the test being administered is sufficiently severe, then we can better understand the ways in which failure to predesignate did and did not matter to CDF's evidence claim.

In Chapter 7 I also discuss a method for evaluating the strength of evidence when a test is suspected of being biased by tuning on the signal and argue that reliable experimental inference requires an adequate model of the error probabilities of the test. Such a model might require taking into account statistically relevant aspects of the testing context involving the behavior and dispositions of experimenters. However, when such social-psychological factors become relevant, an adequate model may be difficult for experimenters to generate. Many rules of experimental method are thus designed to make such factors statistically irrelevant. Hence, evidential relationships do not exist on a plane free of considerations of human sensibilities and preferences.

On the surface, such an acknowledgment may seem to pose a threat to the alleged objectivity of the error statistical concept of evidence. Certainly it has been taken as such by critics of classical statistical testing techniques and by some who have rejected the evidential significance of predesignation. The objection posed is roughly that intentional and epistemic factors cannot be relevant to an objective evaluation of the evidential significance of an experimental outcome. I respond by drawing a distinction between the causal and epistemic relevance of such states. Evaluating experimental evidence in an error statistical manner may require consideration of the causal relevance of experimental agents' intentional or epistemic states. Nonetheless, error statistical evidence is entirely objective in that such states are not epistemically relevant to its evaluation.

The experimenters discussed here worked in a world filled with physical and social complexities but sought to isolate a single and in some sense simple fact in the midst of that tangle. The experimenter interacts with her colleagues and competitors, but also – as an agent with her own beliefs, preferences, and weaknesses – with the objects on which the experiment is performed. This account seeks to elucidate simultaneously the thicket in which the experimenter works and the potential emergence from that thicket of an objective reason for believing some important truth about the world. That such a thing could happen at all is in itself wondrous and seemingly magical, but in truth hard work, careful planning, and fortuitous circumstances make it possible. In any case, that is how things happened in the story that follows.

1

Origins of the Third Generation of Matter

In the *standard model* (see Figure 1.1) of the elementary particles and forces of nature that is endorsed by contemporary physics, the fundamental basis of matter is constituted by the *quarks*, which bear fractional electrical charges and respond to the *strong force* that binds together atomic nuclei, and the *leptons*, which are neutral or have integral electrical charges and are immune to the strong force. The standard model groups these matter particles into three *generations*, where each generation includes two quarks and two leptons. The *up* (u) and *down* (d) quarks, the *electron* (e), and the *electron neutrino* (ν_e) constitute the first generation. The second generation includes the *charm* (c) and *strange* (s) quarks, the *muon* (μ), and the *muon neutrino* (ν_μ). Finally, the *top* (t) and *bottom* (b) quarks join the *tau* (τ) and the *tau neutrino* (ν_τ) in making up the third generation (see Kane 1993 for an introduction).

Another class of particles known as gauge bosons serve as propagators of the elementary forces. These include the massless *photon* (γ, carrier of the electromagnetic force), the *gluon* (g, carrier of the strong force), and the *intermediate vector bosons* W^+, W^-, and Z^0 (carriers of the weak force responsible for various decay processes).

By the early 1990s, all these particles had been experimentally confirmed, except for the top quark. So when the Collider Detector at Fermilab (CDF) collaboration announced in 1994 that they had found the first evidence of the top quark, it was a major scientific development. CDF presented their results in a long, detailed article submitted to *Physical Review D*, entitled "Evidence for Top Quark Production in $\bar{p}p$ Collisions at $\sqrt{s} = 1.8$ TeV" (Abe, Amidei, et al. 1994b). Prominent among the many works that CDF cites in this 60-page article is a 1973 article by Makoto Kobayashi and Toshihide Maskawa (Kobayashi and Maskawa 1973; Figures 1.2 and 1.3).

Citations in a scientific paper perform a variety of tasks, such as referring the reader to other work that supports particular claims being made in the present article and helping the reader to locate articles expressing

		first generation	second generation	third generation
fermions	quarks	***u***	***c***	***t***
		d	***s***	***b***
	leptons	ν_e	ν_μ	ν_τ
		e	μ	τ

gauge bosons
γ
g
Z
$W^\pm$

Hadrons	
Mesons	**Baryons**
$\langle q\bar{q}\rangle$	$\langle qqq\rangle$
e.g., $K^+ = \langle u\bar{s}\rangle$, $J/\psi = \langle c\bar{c}\rangle$	e.g., $p = \langle uud\rangle$, $\Omega^- = \langle sss\rangle$

FIGURE 1.1. The particles of the standard model.

FIGURE 1.2. Toshihide Maskawa. Courtesy of Makoto Kobayashi.

competing viewpoints. Citations may help to validate the competence of the researcher with respect to the subject matter at hand (Kitcher 1995). Many citations, however, are simply meant to acknowledge the contribution that a particular work has made in developing the ideas under discussion. CDF's citation of Kobayashi and Maskawa's article is honorific in this sense. It reflects physicists' consensus that Kobayashi and Maskawa, in their 1973 article, were the first to suggest that something like the top quark might exist.

The citation of the Kobayashi and Maskawa (KM) paper appears in the first sentence of CDF's article, which reads, "The standard model has enjoyed outstanding success in particle physics for two decades, yet one of its key constituents, the top quark, has remained unobserved" (Abe, Amidei,

FIGURE 1.3. Makoto Kobayashi. Courtesy of Robert Palmer, Brookhaven National Laboratory.

et al. 1994b, 2968). Two reference numbers follow the words "The standard model." The first directs the reader to citations of articles by Sheldon Glashow, Steven Weinberg, and Abdus Salam (Glashow 1961; Weinberg 1967; Salam 1968). These citations all point to what are regarded as the seminal works in the development of the standard model of electroweak interactions, which unifies the weak force and the electromagnetic force. The second number appearing after "the standard model" refers to citations of a 1970 article by Glashow, John Iliopoulos, and Luciano Maiani, and to Kobayashi and Maskawa's 1973 paper. The Glashow-Iliopoulos-Maiani (GIM) paper (Glashow, Iliopoulos, and Maiani, 1970) proposed a model of weak interactions based on a quartet of fermions. All the papers that CDF cites in connection with "the standard model" are among the most cited articles in the literature of particle physics.[1]

The KM paper is entitled, "*CP*-Violation in the Renormalizable Theory of Weak Interaction." The abstract reads as follows:

> In a framework of the renormalizable theory of weak interaction, problems of *CP*-violation are studied. It is concluded that no realistic models of *CP*-violation exist in the quartet scheme without introducing any other new fields. Some possible models of *CP*-violation are also discussed. (Kobayashi and Maskawa 1973, 652)

The reader may reasonably be wondering what any of this has to do with the top quark.