Agency and Objectivity in the Search for the Top Quark

Kent W. Staley*

Saint Louis University

March 31, 2004

Abstract: From the perspective of Mayo's error statistical theory of evidence, I explore problems and prospects for an account of the objectivity of scientific evidence. A recent proposal by Peter Achinstein provides the starting point. I consider a challenge to this proposal arising from the role of agents in carrying out the testing procedures that are central to the error statistical theory. Achinstein's objective concept of unrelativized potential evidence initially resolves these difficulties, only to give way to a deeper incompatibility between Achinstein's conception of objectivity of reasons to believe and the error statistical theory of evidence. I propose an alternative account of objectivity of reasons that is compatible with the error statistical theory.

1 Introduction

On April 26, 1994, the Collider Detector at Fermilab Collaboration announced the submission of a paper to *Physical Review D* bearing the title "Evidence for Top Quark Production in $\overline{p}p$ Collisions at $\sqrt{s} = 1.8$ TeV." The word "evidence" reappears in the conclusions at the

*I am grateful to Peter Achinstein for an illuminating and extended correspondence over the issues discussed here, and for urging me to clarify my own views—not just to state my views more clearly, but to adopt views that are hopefully clearer than those I started with. I also benefited from helpful discussions with Dianne Brain and with colleagues at Saint Louis University: Susan Brower-Toland, Alicia Finch, Dan Haybron, Scott Ragland, and Joe Salerno.

end of the paper: "The data presented here give evidence for, but do not firmly establish the existence of, $t\bar{t}$ production in $\overline{p}p$ collisions at $\sqrt{s} = 1.8$ TeV" (Abe et al. 1994, 3023). What sort of a claim is this?

Some philosophers interpret such evidence statements, in some contexts, as statements of objective facts that obtain independently of what anyone does or should believe. Others (personalist Bayesians, for example) take such statements to express something primarily about what some person or persons either do or should believe the facts are, given their other beliefs. One advocate of the former viewpoint is Peter Achinstein.

In *The Book of Evidence*, Achinstein defends the centrality to scientific reasoning and practice of what he calls *potential evidence*. According to his account, if some fact E is potential evidence for a hypothesis H, then E is a good reason to believe H, even if no one knows or even believes that E is evidence for H, even if no one actually believes H for the reason that E, and even if no one knows or even believes that E is the case. This concept of evidence does not require that H is true. According to Achinstein, "E is *veridical evidence* for H" is true just in case E is potential evidence for H, and H is true.[1] Crucial to Achinstein's way of conceiving the objectivity of evidence is that potential evidence is not relativized to any person or group of persons, or to any belief set or epistemic situation.[2] Achinstein is thus committed to the following principle:

[1]Achinstein also defines a stronger concept of veridical evidence that requires, in addition, that there is an explanatory connection between E's being true and H's being true (Achinstein 2001, 174). His full treatment of potential evidence requires only that, in addition to E and background knowledge B both being true and E not entailing H, there be a high probability, given E and B, that there is an explanatory connection between H and E (Achinstein 2001, 170).

[2]Achinstein accepts that some concepts of evidence are thus relativized, but distinguishes them from the more central unrelativized concepts of potential and veridical evidence. *ES-evidence* is relativized to an epistemic situation. If E is ES-evidence for H, then E is a good reason for anyone in a particular epistemic situation to believe H, whether or not anyone is actually in such an epistemic situation, or, if someone is, whether that person does believe E, H, or that E is evidence for H. *Subjective evidence* is relativized to persons. If E is S's subjective evidence for H, then S believes H for the reason that E. It need not be the case that E is a good reason for believing H, even relative to S's epistemic situation.

β: If E is potential evidence that H, then E is a good reason to believe H, independently of any person or epistemic situation.

Another advocate of objectivity in the theory of evidence is Deborah Mayo. In her *Error and the Growth of Experimental Knowledge*, she contrasts the objectivity of her concept of error statistical evidence with the pernicious subjectivity of personalist Bayesianism. Mayo is less systematic than Achinstein in specifying what such objectivity amounts to, but at least two main points are central to her understanding of the objectivity of error statistical evidence. First, evidential judgments are corrigible, and hence are subject to error. She endorses Henry Kyburg's insistence that the possibility of error is a "touchstone of objectivity" (Mayo 1996, 83). Second, evidential relationships supervene on the error probabilities of testing procedures, which are not a matter of opinion, but are facts about long run relative frequencies.

My question is whether the concept of objectivity codified in Achinstein's principle β is compatible with Mayo's error statistical theory of evidence. The motivation for undertaking this investigation is that the error statistical theory strikes me as a promising account of scientific evidence with strong connections to scientific practice. Yet the sense in which it is objective requires clarification. Achinstein's careful articulation of different concepts of evidence with distinct types of objectivity provides a useful starting point for such an investigation, although Achinstein employs these concepts in defense of his own theory of evidence, which I will not here discuss.

I wish to explore two sources of difficulty facing an Achinsteinian understanding of error statistical objectivity. The first concerns the role of the experimental agent in performing the tests that are central to the error statistical theory of evidence. This difficulty will prove to be merely apparent. However, the resolution of the challenge from experimental agency will highlight a deeper problem with β concerning the nature of reasons, and whether they can be the kinds of things Achinstein claims. I will offer in place of β an alternative formulation of the sense in which error statistical evidence provides objective reasons to believe empirical claims.

2 The Error Statistical Theory of Evidence

In Mayo's account, experimental evidence for an empirical claim arises from a process of testing, and the evidential status of the results of experiment depends upon certain characteristics of the testing procedure.

Specifically, suppose that we subject our hypothesis H to a testing procedure T that yields a result E. Then, on Mayo's account, E is evidence for H just in case:

(1) E fits H, and

(2) H passing T with E constitutes a severe test of H.

Satisfying (1) is a matter of degree. The point of requirement (1) is that H and its competitors are not simply a means of deducing some disorganized collection of predictions; rather, some possible results should be "closer," in some — typically probabilistic — sense, to what H leads us to expect than others; E should be a result that is close to what H leads us to expect. Requirement (2) is to be understood as equivalent to the following probability claim:

(2') Prob (H passes T with a result such as E|¬H) is very low,

where "a result such as E" should be understood as "a result that fits H as well as or better than E." If H is being tested against a compound hypothesis—one that comprises more than one distinct alternative to H—then the probability in (2) will not be well-defined, and we must instead evaluate the probability of H passing T with E under specific alternatives to be ruled out. In any case, since "very low" admits of degrees, so does the satisfaction of requirement (2).

The probabilities referred to in this account are intended by Mayo to be objective frequentist probabilities. E being evidence for H depends on the probability distribution for the outcome space (including E) determined by H, and on the probabilities of H's passing T with an outcome such as E determined by the alternatives against which H is tested. Hence, under the error statistical theory, one can make the following claim: *evidential status supervenes on strictly objective facts.*

Though this claim seems straightforward at first, it is worth pausing to probe more deeply the conceptual elements in this account. Some issues will then arise that challenge at least some common conceptions of objectivity.

3 Are Tests Objective? A Problem from the Search for the Top Quark

What counts as evidence on the error statistical analysis is not a fact of just any sort, but a fact regarding the outcome of a testing procedure. Whether E is evidence for H is not a matter of E alone, but of the testing procedure that produced E as an outcome.

Or, to put it another way: In an experiment, data are collected, but these data by themselves are of no use. They cannot be said to be evidence for a hypothesis apart from some specification of the test to which that hypothesis is being exposed by means of these data. In terms that Mayo adapts from Patrick Suppes, a model of the data is needed, as well as models of the experiment and the hypothesis (Suppes 1962). Such models are specified in part through the determination of test parameters, and error probabilities are then evaluated based on those models. These error probabilities then allow one to evaluate whether the test employed by experimenters has severely tested the hypothesis.

The referent of the term "the test employed by experimenters" may be anything but clear, however. I will show next, by returning to the example of the top quark evidence mentioned in the beginning of this essay, that in some contexts the referent is quite indeterminate, and that when it is determinate, it may be determined in part by the intentions of experimenters.

The Fermilab Tevatron is a particle accelerator that collides protons with antiprotons at very high energies. CDF searched for the top quark by surrounding these collisions with a complex, barrel shaped detector designed to measure the properties of particles produced in the collisions and their subsequent decay products. Top quarks, if they existed, would be produced very rarely, in top-antitop ($t\bar{t}$) pairs that would then decay according to certain characteristic patterns, or "signatures." Principally, the top quark decays directly to a W boson and a b quark. But these also decay, and CDF's search for the top quark focused on identifying particles whose ancestry could be traced to a top decay. One important signature would involve a high transverse-momentum lepton (either an electron or a muon), three or more high energy jets of strongly-interacting hadrons, and another electron or muon (produced by the decay of a b quark) with low transverse momentum—a soft lepton. The search for events bearing this signature was called soft lepton tagging (SLT).

"High momentum," "low momentum," and the like are vague terms. CDF sought to make them precise in order to distinguish real top quark decays (signal events) from background processes that might mimic this top quark signature (background events). They did this by choosing the values required of various particle measurements (cuts) to constitute a candidate event. Any collision event that satisfied the cuts would qualify as a candidate event. Having chosen a set of cuts, CDF could then compare the number of candidate events in their data with their estimate of the average number of candidate events they could expect to find in such a data set from background sources alone. The existence of the top quark would manifest itself as a significant excess in the number of candidate events beyond the expected background.

What constitutes a significant excess? Quantitative error statistics can help address this question. CDF had determined for themselves a null hypothesis:

H: This data sample has been drawn from a population of proton-antiproton collision events that is free of top quark production.

They sought to test this against an alternative hypothesis:

J: This data sample has been drawn from a population of proton-antiproton collision events that contains some top quark–producing events.[3]

[3]This constitutes a compound alternative hypothesis, comprising different possible rates of top quark production. That rate (determined by the $t\bar{t}$ production cross section $\sigma_{t\bar{t}}$) is a function of the then unknown top quark rest mass.

For the purpose of such a test, they defined a test statistic:

X $\equiv$ the number of candidate events in the present data sample

With these elements in place, they produced a null probability distribution for X. This distribution gives the probability of getting various values for X, assuming that H is true, for the experiment being performed. From the null distribution they could then produce an estimate of the expected background.

After collecting data for 18 months beginning in 1992, CDF had data on approximately 16 million collision events. Among these, they found seven SLT candidate events. Based on their null probability distribution, they estimated that they should expect on average approximately three SLT candidate events from background.[4]

Given that outcome, they then sought to calculate the significance level of their results—i.e., the probability of getting seven or more candidate events, assuming the null hypothesis H. They found that, were there no top quark, and were they to repeat their experiment infinitely many times, they would get seven or more candidate events about 4% of the time.

CDF had used the SLT search during a previous data-collecting period in which they found no evidence for the top quark. However, having failed to find the top quark, CDF was able to establish a minimum value for its mass (Abe et al. 1992), since theory dictated that the lower the top quark's mass, the more frequently the particle would be produced, and the more quickly it would show up in their data.

As CDF prepared to begin a new round of data-collection in 1992, some discussed the possibility of changing some of the cuts used in the SLT search. The debate focused on the minimum value required for the momentum of the soft lepton. The minimum value had been set at 2 GeV/c. Some argued that the cut should be moved to 4 GeV/c on the grounds that, given a more massive top quark, leptons with momentum in the range from 2–4 GeV/c were

[4]The SLT search was just one element in a complicated experiment. CDF's evidence announcement included other sources of evidence not discussed here. Their estimate of the significance of their combined results was 2.6×10^{-3}, which excluded several sources of information regarded as positive cross-checks on their result. See Staley 2004 for more.

much more likely to come from background than from top quark decays. This argument was not absolutely conclusive, however. The physicists chiefly responsible for the SLT search at the time thought they had good reasons to keep the cut at 2 GeV/c—not least in order to maintain continuity with the earlier search. Two facts about the CDF collaboration at the time of these events set the stage for the controversy that surrounded the SLT analysis. First, CDF members were free to examine new data as it became available. Second, the two physicists who did most of the work on the SLT algorithm worked very independently from the rest of the group.

CDF eventually reported the SLT results with the soft lepton cut at 2 GeV/c. However, some physicists in the collaboration expressed uncertainty regarding both the timing of this choice and the way in which the choice was made. Three of the seven candidate events found by the SLT analysis would be excluded if the cut was moved up to 4 GeV/c, yielding an apparently less significant result. Some collaboration members consequently worried that the apparent significance of the SLT results was an artifact of a manipulation that created the appearance of a genuine effect out of mere background. Particle physicists consider such manipulation—intentional or otherwise—a sufficiently serious problem to merit a special name: "tuning on the signal."[5] In this case, the availability of data for scrutiny entailed that such manipulation was possible, and the relatively private nature of the SLT development process made it easy for other physicists to worry that it had occurred, even if not deliberately.

To understand what is troubling about tuning on the signal, consider the officially quoted significance level for CDF's SLT search: 0.041. Based on the assumptions CDF was making, if the null hypothesis were true, and one were to repeat infinitely many times an experiment using the same detector, using the same cuts, collecting the same amount of data, and so on, one would get as many as seven candidate events or more only 4.1% of the time.

However, if we know that the cuts used in this case were chosen in such a way as

[5]See Franklin 2002 for a discussion of this type of problem as it arose in several experiments in physics.

to exaggerate the apparent significance of the results, then we have statistically relevant information about the experimental procedure used to reach these results. Specifically, the procedure followed—including now the procedure for choosing the cuts—has different error characteristics than the procedure on which CDF based their significance estimate of 0.041. That estimate was based on the specification of a reference class, which is a hypothetical population of repetitions of the testing procedure used in the experiment. If experimenters have tuned their cuts on the signal, then a reference class that would otherwise be appropriate would be the wrong choice for calculating that probability.

In short, the problem that the SLT analysis posed for some in the CDF collaboration was that uncertainty over the procedure used for determining the cuts meant uncertainty over the reference class to consider when evaluating error probabilities. For those who regarded the characteristics of the testing procedure as having been rendered unclear, it became difficult to assess reliably the severity of the testing procedure employed with respect to the top quark hypothesis as a whole (see Staley 2002, 2004).

4 Objective Evidence Supervenient on Actions

If, as the error statistical theory has it, evidence for a hypothesis requires the passing of a severe test by that hypothesis, then one cannot say whether a particular fact constitutes evidence for a particular hypothesis without reference to some test that the hypothesis in question passes with that result. But even if things like quarks exist independently of our beliefs or actions, it does not seem that tests do. Testing procedures exist through the actions of persons who conduct those tests. The CDF data were not by themselves evidence for the top quark. Such status depends on the testing procedure, which depends, not only on the cuts used to define a candidate event but also potentially the decision procedure used to determine those cuts. And certainly that cannot exist apart from actions of experimenters.

The example of the SLT controversy points to a potential feedback loop: the fact that a certain outcome would obtain relative to one set of test criteria can stimulate a change in the criteria by the experimenters. Indeed, the adjustment of test criteria to achieve a certain outcome needs to be recognized as itself part of the procedure being enacted in those cases. When the effect of this feedback loop on the error probabilities of the procedure is not taken into account, estimates of severity become biased and misleading.

Furthermore, in a collaboration these test criteria are determined by a collective decision, and different members of the collaboration may have different views of the procedure being enacted, both in terms of explicit data selection criteria and in terms of a broader understanding of the test procedure. CDF members disagreed as to whether the SLT test procedure included a feedback loop. In such cases, it may not be clear just what constitutes the factuality of any specification of the test procedure employed.

The testing of a hypothesis by a collaboration is a collective action, and thus requires some form of shared intention. But collaboration members may intend to test the hypothesis in different ways. If that is the case, there may not be one test that the hypothesis is subjected to. The outcomes of the multiple tests intended by the different members of the group may have divergent evidential statuses. In practice, this problem is avoided largely through the demands of publication. The need to come to agreement on a single document representing the collaboration's collective understanding of their results does much to bring the requisite shared intention about. Individual collaboration members subordinate their intention to test the hypothesis according to particular criteria to their intention to test it according to criteria agreed upon by the collaboration.

In the debate over the SLT analysis, the demands of publication resulted in an agreement on the final choice of transverse momentum cut at 2 GeV/c, in spite of the fact that some collaboration members remained convinced that it was not an optimal choice. (As one collaboration member put it, "I could live with that" (Contreras 1995).) Absent such agreement, there may be no determinate fact of the matter as to just what test a particular hypothesis is being subjected to, and hence no determinate evidential status for the data.

It cannot be said, then, that CDF's data constitute evidence for the top quark independently of any person's intentions or beliefs. The testing procedure matters, and it is determined in part by the intentions of the experimenters conducting the test. For some testing procedures, the data may indeed yield evidence that there is a top quark, while for others they may not do so.

5 An Achinsteinian Response

Is this compatible with Achinstein's conception of objective evidence? It appears compatible. If one accepts the error statistical theory of evidence, one could nonetheless insist that principle β holds for those test-outcome facts that do serve as evidence.

β_E: If H's passing T with outcome E suffices for E to be evidence for H then E is a good reason to believe H independently of any epistemic state, including whether or not anyone knows or believes (among other things) that T constitutes the procedure used to test H, or that H's passing T with E suffices for E to be evidence for H.

Achinstein endorses a view similar to this for cases where "selection procedures" for collecting data are evidentially relevant (2001, 213n). The outcome of a testing procedure is person-dependent in the sense that it could not occur unless some persons carried out the procedure in question. Nonetheless, that the procedure was carried out, that it had such an outcome, and that it has specific error probabilities are objective facts in the sense that their factuality is independent of our opinions about their factuality. In the kind of case, alluded to above, in which the testing procedure is indeterminate, we would have to say that evidential status is likewise indeterminate. Yet when evidential status is, in error statistical terms, determinate, the facts that determine it exist independently of being believed or known, and principle β_E is meant to reflect this point.

However, Achinstein does not accept that E can only be evidence for H relative to a testing procedure (2001, 214). So the second aspect of his response to these issues will be:

A: Even if there are some cases where we cannot say whether data constitute evidence for a hypothesis without specifying a testing procedure, there are other cases where a

> particular fact just is a good reason to believe a particular hypothesis, independently of any testing procedure.

Whereas on the error statistical theory, evidential status is *always* relative to the testing procedure employed, on Achinstein's account the testing procedure is only sometimes relevant to evidential status.

Proposal β_E is simply an application of principle β to cases where the testing procedure employed is evidentially relevant, and as such *appears* compatible with the error stastistical theory. On the other hand, thesis A, if correct, would call the very terms of the error statistical account into question. But I will argue that the kinds of cases invoked on behalf of thesis A can be easily accounted for on the error statistical account, provided we give up β as Achinstein understands it.

The argument for A is straightforward. In some cases, a fact can be seen obviously to constitute evidence for some hypothesis without any consideration of a test procedure. Therefore it is possible for a fact to constitute a reason to believe a hypothesis regardless of the testing procedure used with regard to that hypothesis. (Being seen to be evidence is not required for the evidence to really be such; the point is rather that it is because we do know of such cases that we have reason to accept the philosophical claim that such cases do exist.) For example, to borrow an example from Achinstein, Sam's spots (having certain specified characteristics) are evidence that Sam has measles. This does not strike us as an incomplete statement lacking a truth value prior to the specification of a testing procedure. The statement is true, with or without a specification of the testing procedure employed, if any.

An initial error statistical reply would be that the most that such examples show is that there are cases in which a fact's status as evidence for a hypothesis is somewhat insensitive to just what test is employed, amongst any of the testing procedures we might regard as likely candidates, not that the testing procedure is irrelevant to whether it is evidence. It is hard, though not impossible, to imagine a testing procedure that passes a hypothesis that a patient has measles when spots of this sort are present that fails to meet the requirements of severe testing. In the cases where it is obvious that an evidential relation obtains, without mention of the testing procedure, we typically imagine that a person confronted with a particular observation (of spots on Sam's skin) just responds to that observation with an inference (that Sam has measles). In such a case, the inferential process itself can be regarded as a test. The inference proceeds according to a certain rule or "precept"[6] such as "Whenever I observe spots with such-and-such characteristics, I infer the presence of measles." On the error statistical account the evidential force of the spots requires that inferences drawn according to this precept, regarded as tests of the hypothesis inferred, satisfy the severity requirement. Suppose, however, that one based one's belief that Sam has measles on the spots Sam has, but reached that conclusion by an inference drawn according to the precept, "Whenever I observe spots with such-and-such characteristics, if I make the observation on Monday or Wednesday I infer the presence of measles; otherwise I infer the presence of kidney stones." In an inference drawn according to such a rule, Sam's spots do not function as a good reason to believe that Sam has measles.

Achinstein will not accept this line of argument, however. He will say that the argument only shows that a fact E might be evidence for, and hence a good reason to believe, a particular hypothesis H, without it being the case that every person who believes H for the reason that E is justified in doing so. But potential evidence is not relativized to any person or to any epistemic situation. Even though the person reasoning according to this bizarre precept would not be justified in concluding that Sam has measles on the basis of Sam's spots, those spots are nonetheless a good reason to believe that Sam has measles, since they are potential evidence of measles. They would be such even if no one were aware of them.

[6] The term "precept" here is drawn from Peirce. See, e.g., his "Theory of Probable Inference" of 1883 (Peirce 1931–1958, 2.694–754, esp. 735).

6 Abstraction or Idealization?

To complete my reply to thesis A, therefore, I must explain the flaw in this concept of "a good reason to believe" employed by Achinstein. In doing so, I will also offer my response to β_E, and in the process suggest a reformulation of β that avoids the difficulties Achinstein's account faces.

Achinstein argues that the principal concern of scientists such as the members of CDF is not simply to demonstrate that their results are *their* evidence for the top quark, or that their results are evidence for the top quark for a person in some particular epistemic situation. Rather, their results are evidence for, and hence a good reason to believe, that the top quark exists, "period." What does this mean? How should we understand principle β?

On the one hand Achinstein holds that the reasons involved in potential evidence claims are not limited in their relevance to just some persons or some epistemic situations: "in this sense, if it is reasonable to believe h, it is reasonable for anyone to do so" (Achinstein 2001, 96). This might suggest that he intends β to be understood in what I will call the "good for everyone" sense:

β': If E is potential evidence that H, then for any person, regardless of epistemic situation (or any epistemic situation, regardless of its justification) E is a good reason for that person (or any person in that epistemic situation) to believe H.

On the other hand, Achinstein holds that these reasons have their status as reasons quite apart from any beliefs actually or potentially held by anyone, which might suggest a "good, but not for anyone" reading:

β'': If E is potential evidence that H, then E is a good reason to believe H, but this does not entail anything about whether E would be a good reason for any person (or any person in a particular epistemic situation) to believe H for the reason E (even if in fact it would be).

To reconcile these two notions, Achinstein draws a distinction between an "abstract" and a "non-abstract" sense of "good reason to believe." The correct interpretation of β is then:

β''': If E is potential evidence that H, then for any person (or any epistemic situation) E is a good reason in the abstract sense for that person (or a person in that epistemic situation) to believe H; but this does not entail anything about whether E would be a good reason in the non-abstract sense for any person (or any person in a particular epistemic situation) to believe H (even if in fact it would be).

To illustrate, let us postulate an experiment just like that performed by CDF and with identical results, but performed by the CDS ("Collider Detector at Sfermilab") collaboration. Suppose that the members of CDS came to believe, with good reason yet incorrectly, that there were severe biases in their algorithms, and that consequently the 12 candidate events they identified are not evidence for the top quark (although in fact they are evidence for the top quark). On Achinstein's account, the CDS physicists would *not* be justified, in the non-abstract sense, should they subsequently believe the top quark to exist for the reason that they had found these 12 candidate events. Given their epistemic situation, they should not adopt such a belief. They *would* be justified in the abstract sense, however, in adopting this belief, because after all those 12 events really are evidence for the top quark.

More generally, according to Achinstein, the aim of scientists in performing experiments is to come to have beliefs about scientific hypotheses that are justified in both senses. That is, they seek to to have evidence for their scientific beliefs that justifies those beliefs both in the non-abstract sense (because the evidence constitutes a reason for anyone in their epistemic situation to believe that hypothesis in question) and in the abstract sense (because the evidence constitutes an abstract reason to believe that hypothesis independently of epistemic situation).

Why should we accept this multiplication of categories of reason? There seem to be two sources of justification, one of which is somewhat holistic and philosophical, and one of which is more empirical and historical. The holistic reason, not argued explicitly by Achinstein, is the role that abstract good reasons play in his overall theory of evidence and objective epistemic probability. Assuming that this theory as a whole is preferable to other theories of evidence and probability, and that abstract good reasons are essential to the theory, such reasons escape Occam's razor and we should accept them into our ontology. Although I think the error statistical theory, which does not require abstract good reasons, has advantages over Achinstein's theory, I do not intend to press this line of argument here.

The second reason, which Achinstein does invoke explicitly, is that such abstract good reasons are necessary to make sense of scientific practice. He pursues this argument through the discussion of episodes such as Hertz's and Thomson's work on cathode rays. The following rough sketch is intended to convey some idea of how this argument is supposed to proceed:

In 1883, Heinrich Hertz tested the hypothesis that cathode rays carry an electric charge by passing them through an electric field within a cathode ray tube. No deflection was observed, and Hertz concluded that his results constituted evidence that cathode rays are not electrically charged. J. J. Thomson later suspected that Hertz's experiment was flawed due to residual gas in the tube. He speculated that the charged cathode rays may have ionized the gas in the tube, and that the ionized gases subsequently shielded the cathode rays from the electric field Hertz had introduced. Using superior vacuum technology, Thomson repeated Hertz's experiment in 1897 and observed the deflection that Hertz missed. Thomson then went on to perform further experiments to show that the "corpuscles" constituting cathode rays are much smaller than hydrogen atoms.

Achinstein's interpretation of this episode is that Thomson concluded, not that Hertz's results had been evidence against charged cathode rays but now were not, but that they never were. Although Hertz's results seemed to provide a good reason to believe that cathode rays are not electrically charged, they did not in fact do so (except relative to Hertz's epistemic situation), due to the flaw in Hertz's experiment. Hertz's evidence claim was refuted by Thomson, and replaced by evidence providing a good reason to believe that cathode rays carry a negative charge:

> Thomson could have provided a justification of belief in the electrical neutrality

> hypothesis simply by citing the results of Hertz's experiments and his own preliminary results. That would have sufficed to justify a belief in the neutrality hypothesis on the part of those physicists (up to 1897) in roughly Hertz's epistemic situation. Thomson wanted to do something more powerful, viz. to provide a good reason to believe that cathode rays are negatively charged—a reason not tied to any actual or hypothetical epistemic situation. (Achinstein 2001, 34–35)

Achinstein concludes that veridical evidence, which requires the existence of abstract good reasons, is in fact the goal of scientific investigation:

> A scientist wants to know whether some experimental results reported in e provide a good reason for believing a hypothesis h—not a good reason for someone in some particular epistemic situation, and not just a good reason for him, but a good reason period, independent of epistemic situations. (Achinstein 2001, 37)

Achinstein and I agree that Thomson sought reasons for believing in the charged cathode ray hypothesis that would survive improvements in the epistemic situations of his fellow and future physicists—something that Hertz's reasons for believing in the electrical neutrality of cathode rays failed to do. Achinstein believes that he also wanted something more than this, but I do not see the evidence for such a claim.

What is apparent from episodes such as the Hertz/Thomson experiments and the search for the top quark is that scientists in making evidence claims are concerned not to be mistaken—that is, they regard evidence claims as corrigible statements of fact that have their factual status independently of any person's beliefs regarding them. It does not follow that Achinstein's abstract good reasons are the only, or the best, way to make sense of this objectivity.

Are there reasons for resisting the claim that there is an abstract sense of justification that calls for the existence of abstract reasons unrelated to any epistemic situation? I think there are, but I confess that they are "merely philosophical." I will try to explain my reservations in what follows.

The claim that some facts constitute reasons to believe certain claims only in the abstract sense serves to resolve the tension between the "good for everyone" and "good, but not for anyone" readings mentioned above. Contrary to the "good, but not for anyone" reading, potential evidence E for hypothesis H constitutes a reason for everyone, regardless of epistemic situation, to believe H. Contrary to the "good for everyone" reading, it does not do this in a sense that entitles us to say that everyone would necessarily be rational in believing H for the reason E. After all, the hypothetical CDS collaboration would not be rational to believe the top quark hypothesis on the basis of their own data, since they believe their own analysis to be so badly biased as to defeat any such evidence claim based on their data.

Achinstein's claim about the aims of science then comes to this: Scientists seek evidence not only because they are interested in being rational in the formation of their scientific beliefs but also because they are interested in (abstract) reasons themselves. This means that, although a scientist seeks to be both abstractly *and* concretely rational, the knowledge of reasons qua reasons, quite apart from their role in making one rational in one's beliefs, has independent value.

I doubt whether this claim is true, and I intend to show how one can have a robust conception of the objectivity of scientific evidence without being committed to it.

My own view is that there is a strong connection between the concept of reasons and the concept of rationality. Reasons matter to us because rationality matters. Philosophically, we begin with an idea that there is a difference between arbitrary, unmotivated, or poorly motivated personal beliefs or actions and personal beliefs or actions that are rational—those of which we can say, "Yes, it makes sense to me that a person in that situation should reach such a conclusion." Reasons are brought in to make sense of this apparent difference. Wrenched off of these moorings in the rationality of personal beliefs or actions, I find it hard to see of what interest reasons could be.

To sharpen this claim, consider a thought experiment: Suppose a scientist were to be offered a mephistophelean bargain by an evil genius: She would be granted the knowledge of some fact (E) that is evidence for a superstring theory (H), and be made to believe H for the reason E, but only on the condition that no one, not even she herself, would ever be in an epistemic situation such that it would be rational to believe H for the reason that E. I think she should reject this deal. Although she may be interested, as Achinstein has it, in more than a good reason for just some person or just some epistemic situation, it does not follow that she is at all interested in a reason that leaves her just as irrational in her beliefs (in the concrete sense) as if she had adopted them after consulting a ouija board.

I think that Achinstein, on the other hand, is committed to saying that she should at least consider accepting the offer.[7] In coming to know of E, and on the basis of E believing hypothesis H, she can believe H for the reason that E. Even if she is irrational in believing thus in the non-abstract sense (for under the terms of the bargain, she does not even know that E is evidence that H), she will still be justified in the abstract sense of believing H for what is, abstractly, a good reason.

I agree with Achinstein that scientists are not, qua scientists, so self-absorbed that their own rationality is their only concern. The members of the CDF collaboration were intensely interested in the top quark itself, not to mention all of the other exciting physics involved in their experimental pursuits, or the thrill of building one of the most complicated experimental instruments ever devised and making it work. Above all, they wanted to "get it right."[8] What I doubt is whether, in addition to being directly interested in top quarks and bottom quarks and silicon vertex detectors, the elimination of sources of error, etc., they were also interested in abstract reasons for the top quark hypothesis. Yes, they were

[7]Presumably, her actual decision would depend on her assessment of the probability that she might become aware of such evidence on her own, thus achieving *both* concrete and abstract justification. In the case of evidence for supersymmetric strings, this probability should probably be considered quite small. As noted previously, Achinstein holds that scientists want both abstract good reasons and non-abstractly rational beliefs. The question here is whether the former would be of any interest *on their own*. I am troubled, for the sake of both Achinstein's claim and my attempt to argue against it, by the fact that such an outlandish thought experiment seems necessary in order to test our intuitions about how to answer this question.

[8]This phrase I am stealing from correspondence with Peter Achinstein.

also interested in being correct in their claim that their data contained evidence for the top quark, but the disagreement here is precisely over what that interest amounts to.

I do not pretend to have an argument disproving Achinstein's account of what such interest in "getting it right" amounts to. However, I do intend to propose an alternative that may account just as well for the aspects of scientific practice that Achinstein highlights, that fits with a promising theory of evidence (the error statistical theory), and that retains a single category of reasons tied intrinsically to the idea of rational belief.

I propose that what is needed here is not abstraction but idealization. I agree with Achinstein that when scientists go in search of evidence they do not merely seek a reason for a particular person or group to believe the hypothesis that evidence supports. Neither would a scientist be satisfied if his evidence merely provided a reason to believe that hypothesis for just some epistemic situation—or rather, for just any epistemic situation. It does not follow that scientists seek reasons that exist apart from any epistemic situation whatsoever.

Instead, I propose that the virtues of Achinstein's account can be retained without the problematic appeal to abstract rationality if we take objective evidence to provide reasons that are relative to ideal epistemic situations. Scientists go to great length to rule out errors in their understanding of their own experiments. This, I propose, is not due to an interest in reasons that make no reference to epistemic situations at all, but instead reflects a concern that future epistemic situations may include knowledge that defeats their evidence claims. I now turn to articulating this proposal.

When an experiment yields evidence E for for a hypothesis H, this evidence does, as Achinstein claims, constitute a reason to believe that H is true. For whom does it constitute such a reason? My proposal is that it does so at least for anyone with a correct understanding of the experimental conditions – anyone in what I will call an "ideal epistemic situation." The next task is to give an account of ideal epistemic situations that is both substantive and defensible.

Experimenters seek to establish evidence claims that will survive any corrections that future inquiry would reveal, even if pursued indefinitely.[9] In "How to Make Our Ideas Clear," C. S. Peirce famously held that the truth regarding any matter could be understood in terms of the long-run consensus, subject to no further corrections, that would ultimately be reached by a community of inquirers who persisted in investigating that matter (Peirce 1931-1958, 5.388–410; see Misak 1991 for a perceptive defense). Although I will not go so far, I think we can employ a similar idea here. At the center of Peirce's approach is the belief that there is an objective fact as to what an extended inquiry into any matter would reveal, if that inquiry were pursued, even if no such inquiry occurs.[10] Suppose that H is a hypothesis, E is an experimental result that has been produced, and T is a testing procedure that resulted in E. I will call S's epistemic state *ideal* with respect to H, E, and T just in case S knows everything that an indefinitely extended and detailed investigation would reveal concerning whether E fits H, and whether H's passing T with E is the passing of a severe test.

This allows the formulation of a rather different claim about evidence and reasons to believe:

γ: If H's passing T with outcome E suffices for E to be evidence for H, then E provides a good reason to believe H for anyone whose epistemic situation is ideal with respect to H, T, and E.

Some points should be noted immediately. First, neither fit nor severity depends on whether H is in fact true, so that the truth of H is not among the relevant facts included in an ideal epistemic state with respect to H, T, and E. Second, this requirement that E provide a good reason to believe H for anyone in an ideal epistemic state is stringent in the sense that

[9] I do not say that this is their only interest. Achinstein claims that scientists seek veridical evidence, which entails that they seek evidence in support of hypotheses that are in fact true. I am inclined to agree with this latter claim, but I wish to separate the question of whether scientists seek true theories from the question of whether they seek reasons to believe those theories that exist apart from any epistemic situation whatsoever.

[10] It is therefore important to note that Peirce's account is not offered as an irrealist dissolution of truth. To the contrary, Peirce was offering a pragmatic theory about truth that he regarded as resting upon a quite robust realism.

it requires *more* than that the evidence claim should survive all actual future inquiry (which might after all end sooner rather than later in some catastrophe). It should be capable of surviving future inquiry as it would develop supposing it to be extended indefinitely. Finally, I am not asserting that the truth of a hypothesis involves nothing more than that no future inquiry would defeat the evidence claims one has made in support of it, nor that the aim of scientists is only for their hypotheses to be empirically adequate. When an experiment generates evidence for a hypothesis, this constitutes a reason to believe that the hypothesis is true, not merely empirically adequate.

Some may balk at my attempt to replace abstract reasons with reasons relativized to ideal epistemic situations because the latter achieves objectivity of evidence at the expense of requiring the objectivity of counterfactual conditionals. Scientific reasoning, however, is replete with consideration of counterfactual conditionals. In any case, the objectivity of such claims is something that the error statistical theory is committed to already. The commitment implied in γ is simply to the claim that there is a certain class of facts concerning the conditions under which the experiment was performed that are (1) relevant to the evidential judgment at hand and (2) in principle ascertainable, though some may not have been actually ascertained. This class cannot be specified in advance, but will depend on the peculiarities of the particular experiment. It is the job of a good experimenter to try to figure out what they are.[11]

When CDF announced having found evidence for the existence of the top quark, they accompanied their announcement with a detailed argument documenting only part of the exhaustive reasoning and testing employed in justifying their claim. They recounted numerous cross-checks, calibrations, and robustness arguments to establish that their epistemic situation, if not quite ideal, was good enough to make their claim without great fear of being embarrassed by future improvements in their own or others' epistemic situations with regard to their experiment. Skeptics within CDF who worried about biases in the SLT search and

[11]Sometimes even good experimenters fail. When such a failure is pointed out, the good experimenter might say, "If only I'd thought to check that!"

elsewhere worried that their epistemic situation was in fact rather far from ideal, and that the collaboration was at significant risk of being embarrassed. Whether satisfied or concerned, the relevant viewpoint seems to have been not the view from nowhere, but the view from possible epistemic situations more complete than their own.

7 Conclusion

Achinstein is committed to thesis A, which is incompatible with the error statistical theory. His defense of A rests ultimately on principle β'''. I do not claim to have refuted Achinstein's β''', much less to have established my own alternative γ. I do, however, claim to have highlighted the conceptual underpinnings of Achinstein's account, bringing into focus its strong dependence on the concept of abstract reasons. In principle γ I have proposed an alternative understanding of the objectivity of evidence claims that does not depend on abstract reasons. I suggest that this alternative proposal does as well as Achinstein's at taking into account the features of scientific practice that motivate his claims, although a full argument for this claim would go beyond the limits of this paper. Principle γ is compatible with the error statistical theory of evidence. I believe that γ has advantages in terms of ontological parsimony as well as an apparently closer connection to scientific practice, though a full defense of it remains to be articulated.

8 References

Abe, F., M. G. Albrow, et al. [CDF] (1992) "Limit on the Top-Quark Mass from Proton-Antiproton Collisions at $\sqrt{s} = 1.8$ TeV." *Physical Review D* 45: 3921–48.

— (1994) "Evidence for Top Quark Production in $\bar{p}p$ Collisions at $\sqrt{s} = 1.8$ TeV." *Physical Review D* 50: 2966–3026.

Achinstein, P. (2001) *The Book of Evidence.* New York, Oxford University Press.

Contreras, M. (1995) Oral History Interview by K. Staley. Tape Recording. October 17,

1995. University of Chicago.

Franklin, A. (2002) *Selectivity and Discord: Two Problems of Experiment.* Pittsburgh, University of Pittsburgh Press.

Mayo, D. (1996) *Error and the Growth of Experimental Knowledge.* Chicago, University of Chicago Press.

Misak, C. J. (1991) *Truth and the End of Inquiry: A Peircean Account of Truth.* New York, Oxford University Press.

Peirce, C. S. (1931-1958) *Collected Papers of Charles Sanders Peirce*, 8 vols. Eds. C. Hartshorne and P. Weiss. Cambridge, Massachusetts, Harvard University Press.

Staley, K. W. (2002) "What Experiment Did We Just Do? Counterfactual Error Statistics and Uncertainties about the Reference Class," *Philosophy of Science* 69: 279–99.

— (2004) *The Evidence for the Top Quark: Objectivity and Bias in Collaborative Experimentation.* New York, Cambridge University Press.

Suppes, P. (1962) "Models of Data." In *Logic, Methodology and Philosophy of Science: Proceedings of the 1960 International Congress.* Eds. E. Nagel, P. Suppes, and A. Tarski. Stanford, Stanford University Press, 252–61.