Higgs Discovery and the Look Elsewhere Effect

Richard Dawid*†

The discovery of the Higgs particle required a signal of 5σ significance. The rigid application of that condition is a convention that disregards more specific aspects of the given experiment. In particular, it does not account for the characteristics of the look elsewhere effect in the individual experimental context. The paper relates this aspect of data analysis to the question as to what extent theoretical reasoning should be admitted to play a role in the assessment of the significance of empirical data.

1. Introduction. In July 2012, the European Organization for Nuclear Research (known as CERN) announced the discovery of a scalar particle that was likely to be the Higgs particle, by then the last unconfirmed prediction of the standard model of particle physics (ATLAS 2012; CMS 2012). Based on further improvements of the data, this discovery has by now been acknowledged as the discovery of a Higgs particle.[1]

Already in December 2011, CERN had first announced indications for a Higgs particle at both Large Hadron Collider (LHC) experiments, ATLAS and the Compact Muon Solenoid (CMS), with a combined significance of nearly 4σ. Since the agreed-upon statistical limit for acknowledging a discovery in particle physics was 5σ, those data amounted to significant evidence for a new scalar particle but did not constitute a discovery. The present

Received April 2013; revised April 2014.

*To contact the author, please write to: Ludwig Maximilian University Munich Center for Mathematical Philosophy, Geschwister Scholl Platz 1, 80539 Munich, Germany; e-mail: richard.dawid@univie.ac.at.

†I am very grateful to Bob Cousins, Luca Moretti, and a number of undisclosed reviewers for very helpful comments and suggestions. The work was in part financed by the Austrian Research Fund (FWF) project no. P22811-G17.

1. It remains to be seen whether the discovered particle has the properties predicted by the standard model of particle physics or must be understood in terms of an extension of the standard model, such as supersymmetry.

Philosophy of Science, 82 (January 2015) pp. 76–96. 0031-8248/2015/8201-0008$10.00
Copyright 2015 by the Philosophy of Science Association. All rights reserved.

https://doi.org/10.1086/679179 Published online by Cambridge University Press

paper focuses on the epistemic status of the Higgs particle between December 2011 and July 2012. Assessments of the December 2011 data could be divided into two clearly discernible 'camps'. Those who adhered to the letter of the definition of a discovery in particle physics recommended caution and warned against being overly confident based on insufficient data. Others, however, emphasized the striking coherence of the data with theoretical knowledge about the Higgs particle and, on that basis, argued that the evidence for the Higgs was stronger than suggested by the formal statistical analysis. This disagreement reflected a substantial conceptual difference of opinion regarding the overall take on empirical confirmation: it hinged on the question of how to evaluate the 'look elsewhere effect' in the given case.[2] The present article analyzes this conceptual disagreement, which points toward a more general question about the epistemological status of measurement in high-energy physics (HEP) and may be expected to be of increasing importance in the future.

After a brief characterization of the Higgs particle and the look elsewhere effect in sections 2 and 3, the core problem associated with interpreting the data is presented in section 4 by defining two different perspectives on the status of measurements in HEP. This discussion is followed by a brief assessment of the significance of the ensuing analysis in section 5. Section 6 construes the two perspectives in terms of the distinction between a frequentist and a Bayesian understanding of statistical analysis. Section 7 then proceeds to discuss a thought experiment that supports the introduction of Bayesian elements of reasoning in the given case.

2. The Higgs Particle and Its Discovery. The standard model of particle physics introduces the Higgs mechanism in order to explain particle masses.[3] While the conceptual success of the standard model crucially relies on the gauge symmetric structure of the theory, unbroken gauge symmetry does not allow for the observed massive vector bosons and fermionic mass spectra. It was therefore suggested in the 1960s that the electroweak gauge symmetry should be broken at the theory's ground state while the theory's Lagrangian remained gauge symmetric. This structural feature, which is called 'spontaneous symmetry breaking', can be obtained by introducing a scalar field, the Higgs field, with a specific potential. After the Higgs mech-

2. Another aspect of the debate was the understanding that a mistaken announcement of a Higgs particle discovery would have had very serious repercussions for the public standing of high-energy physics. This pragmatic reason for being particularly cautious in the given case was independent from genuine epistemic considerations and will not be addressed in the present paper.

3. The Higgs mechanism was proposed in Higgs (1964) and Englert and Brout (1964). For a recent survey of the historical background, see Karaca (2013).

https://doi.org/10.1086/679179 Published online by Cambridge University Press

anism had emerged as the only promising mechanism capable of reconciling gauge field theory with the observed mass spectrum, the striking conceptual and empirical success of the standard model established increasing trust in the viability of the Higgs mechanism. Once all other predictions of the standard model had been empirically confirmed by the mid-1990s, the Higgs particle remained the only standard model particle to be searched for. The LHC at CERN was built with the aim to test the entire parameter space where the Higgs particle could be expected to be found (apart from a few patches that would constitute particularly unfortunate fine-tuning).

Higgs particles can be produced in deep inelastic scattering processes: particles are smashed together with very high kinetic energy so that, according to the laws of relativity and quantum physics, new kinds of particles are generated in the collision process. If the Higgs particle exists and particle collisions set free energies above the rest mass energy of the Higgs particle, Higgs particles are generated in accordance with the physical conservation laws. The rate of the Higgs production in collisions at a specific energy depends on the involved coupling constants and the phase space (the space of locations and momenta) of the possible experimental outcomes that involve a Higgs particle. The empirical confirmation of the Higgs particle in collider experiments is a particularly difficult enterprise owing to the specific properties of the particle. Because of its high mass, the Higgs particle decays into energetically favored lighter particles after a very short period of time. This time period is too short for generating an observable trace or gap (the standard model Higgs particle is electrically neutral and therefore could not generate a particle trace) in a detector. Generation and decay of the Higgs particle occur, for all practical purposes, at the same spot in the detector and must be attributed to one and the same vertex in a picture of particle traces extracted from the detector.

This fact makes it very difficult to identify Higgs particles on pictures of scattering events. The generation and decay of a Higgs particle can never be univocally attributed to an individual vertex in a scattering picture. Any vertex that can be interpreted as containing a Higgs particle allows for a number of other interpretations that do not involve a Higgs particle. For that reason, the existence of the Higgs particle must be demonstrated on a statistical basis. The empirical analysis consists of two separate parts. First, the rates of specific types of events that contain vertices that might arise as a result of Higgs production must be calculated based on the known and well-established theories of particle physics without assuming the existence of the Higgs particle. This calculated event rate is called the background of the process under investigation. It is then checked whether the collected data are compatible with the calculations of the expected background. This corresponds to testing the null hypothesis that assumes standard model physics without a Higgs particle. If the null hypothesis is excluded, physicists are justified to claim that they have observed new physics. The sec-

https://doi.org/10.1086/679179 Published online by Cambridge University Press

ond step then consists in analyzing whether the data, based on our theoretical knowledge, are univocally consistent with the Higgs hypothesis. If that is established, the collected data can be acknowledged as conclusive empirical evidence for a Higgs particle.

3. The Look Elsewhere Effect. We now want to take a closer look at the first step of the data analysis described above. It deals with the question of whether or not the collected data imply new physics. Let us assume that physicists have calculated the background of the process under investigation. Statistical fluctuations due to the quantum nature of the process can produce events in excess of that expected background. Faced with a specific measured excess rate of events at a certain energy scale, experimentalists thus must calculate the probability that a number of events at least as high as the measured number would be produced at that energy scale if the null hypothesis (asserting the viability of standard model physics without a Higgs particle) were true. That probability is called the local p-value of the data with respect to the given null hypothesis. If the local p-value lies below a certain limit, the null hypothesis can be taken to be refuted and experimentalists are justified in speaking of the discovery of new physics. Particle physicists have set the limit for announcing a discovery at a 5σ confidence level, which—expressed in terms of the Higgs search—corresponds to the condition that the probability of finding an excess rate of at least the measured size that is caused by standard model particles without a Higgs particle is lower than 3×10^{-7}.

Setting a discovery limit is, of course, a matter of convention. It is based on a trade-off between the advantage of calling a viable scientific claim empirically well established and the potential damage of endorsing a false scientific claim. Setting discovery limits somewhere is necessary in order to be able to treat scientific statements as stable elements of a scientific worldview. Without defining any limits, physicists could never claim to have discovered a new phenomenon in microphysics and could not define a univocal conceptual basis for the null hypothesis in future high-energy experiments.

In many scientific fields, a 3σ effect, which corresponds to a local p-value of 0.15%, is considered sufficient for inferring a new phenomenon. In particle physics, the discovery of a new particle is taken to be of very high importance and is used in the analysis of the background in all future high-energy scattering experiments. Therefore, the risk of erroneously acknowledging a discovery of a particle should be kept particularly low, and a stronger criterion seems advisable.

Still, a 5σ limit for discovery might seem surprisingly high at first glance. One important reason for this remarkably high limit has to do with a characteristic aspect of experimentation in HEP that is called the look elsewhere effect (LEE). Normally, experiments in HEP do not just search for new

https://doi.org/10.1086/679179 Published online by Cambridge University Press

phenomena at one specific energy scale but test a wide energy spectrum. The chances that an experiment shows a certain deviation from the predicted event rate at some energy scale within the tested energy range thus must be calculated by summing up the chances of finding it at each specific energy scale. The number of 'places' where one can find a signal is roughly given by the tested energy range over the width of the signal.

To give a specific example, let us imagine that energy levels of specific events can be specified with the accuracy of 1 GeV and a range of 100 GeV is tested. The probability of finding a deviation from the predicted event rates above the 5σ limit somewhere within the tested energy range then is roughly 100 times the probability of finding it at one specific energy level. That is, the probability of getting a 5σ effect owing to statistical fluctuations anywhere within the measured energy range is not 3×10^{-7} but rather 3×10^{-5}. The latter number, which represents the probability of finding an excess rate beyond a given significance level somewhere in a chosen range of testing, is called the global *p*-value of the data. It is the more telling number than the local *p*-value with regard to the question of whether or not one should reject a null hypothesis and therefore infer new physics. In our example, a global *p*-value of 3×10^{-5} may be taken to be a reasonable limit for acknowledging a discovery of a new particle once one takes into consideration that (1) in the face of a considerable number of experimental tests of various kinds of new physics, one wants to have a small probability that any announcement of a discovery is spurious; and (2) it seems wise to introduce some extra error margin in order to account for unknown systematic errors that might distort the empirical data.

The LEE thus is one main reason for setting the significance limit for acknowledging a discovery as high as 5σ. Historically, the 5σ limit was established based on largely pragmatic considerations. While statistical fluctuations at a 4σ level did and do occur from time to time in HEP experiments, no 5σ signal in a particle experiment has, up to this point, ever turned out to be a fluctuation.[4] A 5σ limit therefore seemed plausible simply based on the historical record.[5] The fact that 4σ fluctuations do occur can be statistically explained based on the number of experiments that are carried out in conjunction with the size of the LEEs that usually apply in those contexts.

Let us now look specifically at the Higgs detection at the LHC. Higgs particles can be found at the LHC in a number of different scattering pro-

4. This does not mean that effects that were classified as stronger than 5σ at the time of measurement never evaporated later on. In some cases, such effects later turned out to be attributable to phenomena already known. In other cases, it turned out later on that the assessment of the effect's significance had been fallacious. See Franklin (2013) for an analysis of such cases.

5. For a survey of the early history of the 5σ criterion, see Cousins (2013). For recent takes on the 5σ criterion, see Feldman (2006) and Lyons (2013b).

https://doi.org/10.1086/679179 Published online by Cambridge University Press

cesses. Two types of processes are of particular importance because they allow for a rather precise specification of the Higgs mass and have a comparably small background in conjunction with a sufficiently high Higgs production rate. In the first type of process, a Higgs particle is produced and decays, via a few intermediate steps, into a photon pair. In the second type, the Higgs ends up decaying into four leptons. Both types of Higgs candidate events can be detected at the ATLAS and the CMS detector. When CERN announced its results in December 2011, excess rates of both events had been measured by both detectors. The significance of the entire excess rate of both event types over both detectors was assessed to be somewhere close to 4σ.[6] This corresponded to a probability close to 3×10^{-5} that the observed number of events would be generated as a fluctuation of standard model physics without a Higgs particle. According to a conservative assessment of the energy range where a Higgs-like particle could be detected in the experiment, a Higgs particle was looked for in about 80 energy bins at the LHC. One thus gets a global p-value of about 2.5×10^{-3}. This probability does not even amount to a 3σ effect and therefore was clearly insufficient for establishing the existence of a new particle. The data of July 2012 then had a significance above the 5σ level. This amounted to a global p-value of less than 3×10^{-5} and was sufficient for declaring the data a discovery of a new particle (see fig. 1).

4. Two Ways of Interpreting the Data. The problem to be discussed in this paper can be seen clearly when looking at the situation between the first CERN announcement of a potential Higgs finding in December 2011 and the announcement of a discovery of a scalar particle in July 2012. The official CERN announcement of December 2011 strictly adhered to the 5σ rule and called the data an indication of a possible Higgs particle that did not constitute a discovery. Some commentators emphasized the preliminary character of the measured effect by pointing out that 4σ effects had turned out to be statistical fluctuations on several occasions in the past. The understanding that lay behind the described kind of reasoning shall be called the "experimentalist's position," for reasons that will become clear later on.

Despite the official caveat, however, many particle physicists felt pretty sure already after the announcement of the December data that actual Higgs events had been detected. Some of them did present an argument for their

6. The calculation of the overall significance of data collected in several different experiments is difficult and suffers from ambiguities. See Cousins (2007) for an analysis and further references. It was particularly tricky in the given case owing to the fact that the characteristic energies of the measured excess rates did not precisely coincide. We need not go into the details of this analysis, however. For the present purposes it is sufficient to have an approximate value for the overall significance.

https://doi.org/10.1086/679179 Published online by Cambridge University Press

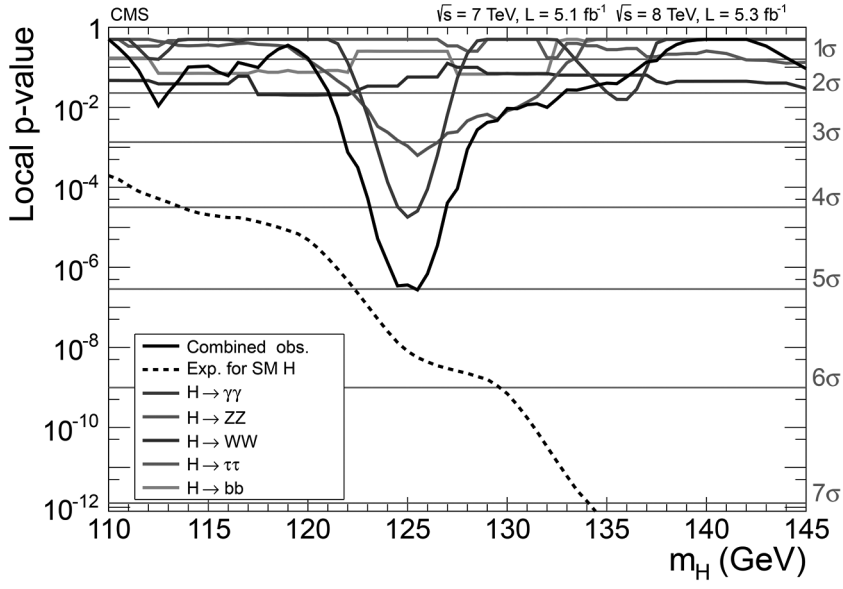


Figure 1. Diagram from CMS (2012) showing the observed local *p*-values for the five decay modes tested at the CMS experiment (*gray-scale lines*) and the overall combination as a function of standard model Higgs mass (*black line*, reaching 5σ at 125.5 GeV). The dashed line shows the expected local *p*-values for a standard model Higgs boson with a mass m_H. Color version available as an online enhancement.

enhanced trust in the significance of the data. They argued that the Higgs case differed from those contexts where 4σ effects had vanished after further experimental testing in the past. In the latter cases, a strong LEE had to be taken into account. In the Higgs case, to the contrary, one could considerably reduce the LEE because (i) one could be quite confident for theoretical reasons that the Higgs existed and (ii) empirically based arguments had already predicted within which energy range the Higgs particle should be found. Physicists who took up that position implicitly suggested that one should look carefully at the theoretical arguments that could constrain the LEE and, on that basis, determine the trustworthiness of the data beyond the rigid application of the 5σ rule. I want to call this position the "theoretician's position," as opposed to the experimentalist's position presented before.[7]

In order to fully understand the difference between the experimentalist's and the theoretician's position, we have to remember the two steps of

7. The 'unofficial' character of the described debate makes it difficult to provide sources for the theoretician's position. The debate is reflected in some of the physics blogs that

https://doi.org/10.1086/679179 Published online by Cambridge University Press

the data analysis mentioned in section 2. The first step establishes that new physics has been found by demonstrating that the collected data cannot be accounted for by the background alone. The second step then must demonstrate that the observed new physics can indeed be identified with the Higgs particle. The close to 4σ effect measured up to December 2011 characterizes the analysis at step one. It can be specified without any knowledge about the Higgs particle. Keeping this part of the analysis free from reasoning based on the Higgs hypothesis is, in fact, based on a core principle of the experimental method: experimentalists want to keep the data analysis as independent as possible from the scientific concepts the data are supposed to confirm (see, e.g., Cousins 1995, sec. 7). If one wants to establish that a given set of data is incompatible with the well-established physics without the Higgs particle, it would feel like begging the question to demonstrate that fact by relying on theoretical reasoning that assumed the Higgs particle's existence.[8] Still, that is what the theoretician's position proposes. According to the theoretician's position, it would be misleading to pretend that we did not have theoretical knowledge that makes us expect the existence of the Higgs particle. By discarding theory-based indications to that end, we may be experimentally flawless but do not provide a realistic picture of the way the collected data actually influence our trust in the Higgs particle.

The designations of "experimentalist's position" and "theoretician's position" are not meant to imply that all or most experimentalists adopt the former and most theoreticians the latter position. Rather, it points at the positions' core concerns. The experimentalist's position has the priority to defend the purity of the process of data analysis by keeping it free of theoretical reasoning that is about to be tested by those very data. The theoretician's position, to the contrary, has the priority to be frank about our actual beliefs with respect to the hypothesis in question. That belief, however, clearly does rely to a considerable extent on the theoretical knowledge we have about the overall situation.

---

extensively commented on the Higgs search. In some entries of http://takingupspacetime.wordpress.com/philosophers-of-physics-the-websites/, the CERN experimentalist Mauro Dorigo defends a clear form of the theoretician's position.

8. The accusation of 'begging the question' must be made more precise. Formally, attributing a prior probability lower than 1 to the hypothesis to be tested is a perfectly valid (Bayesian) move. Since no one pretended to be absolutely sure about the existence of the Higgs particle before its discovery, the claim that the theoretician's position amounts to begging the question thus cannot imply that its line of reasoning is formally invalid. Rather, it suggests that its line of reasoning is "dialectically ineffective" (I am grateful to Luca Moretti for emphasizing this point and providing the term): to those who do not share the trust in the Higgs particle, the reduction of the LEE is not plausible and the argument based on it cannot be convincing.

https://doi.org/10.1086/679179 Published online by Cambridge University Press

Before assessing the strengths and weaknesses of the two positions, we have to specify a little more clearly what the theoretical support for the Higgs data amounted to. This support was based on the understanding that a Higgs particle was likely to exist. When the LHC search for the Higgs particle began, the standard model was an empirically well confirmed theory except for the Higgs particle. Based on that fact, a seemingly cogent line of reasoning generated trust in the existence of a Higgs particle:

1. It was considered highly unlikely that an empirically equally successful description of the physics described by the standard model could be formulated that was entirely independent from the principles of gauge field theory that stood behind the standard model.
2. In order to make gauge field theory compatible with the data, it seemed inevitable to introduce a concept of spontaneous symmetry breaking.
3. The only satisfactory way of introducing spontaneous symmetry breaking (disregarding the far more difficult and, in some respects, conceptually doubtful approach of 'Higgsless models' that were discussed in recent years) seemed to be the introduction of a Higgs scalar—which could be an elementary particle or a bound state.

On the described basis, high-energy physicists felt highly confident about the existence of a Higgs scalar even in the absence of direct empirical evidence for it.

Now, it is important to understand that the Higgs field can have empirical effects in two ways. On the one hand, a Higgs particle can be produced in collisions as a real particle if the collision energy is high enough for generating a particle of its mass. This is the effect searched for at the LHC experiments. In addition, however, the Higgs field has an effect on perturbative calculations of processes that do not generate real Higgs particles. Given the uncertainty principle of quantum physics, calculations of perturbative corrections in quantum field theory must take into account contributions of 'virtual particles', that is, particles that could not be produced in accordance with the laws of energy and momentum conservation. These virtual Higgs contributions to perturbative corrections play a role already at energy levels that are too low for generating real Higgs particles. The size of those effects of the Higgs field on scattering cross sections can be calculated. Calculations of scattering amplitudes that take into account virtual contributions of a Higgs particle of a given mass then can be compared with actual measurements of the corresponding scattering processes at the given energies.

No high-energy experiment before the start of the LHC provided any data that could be understood as an effect of virtual Higgs contributions.

https://doi.org/10.1086/679179 Published online by Cambridge University Press

This observed lack of Higgs-induced effects constrained the Higgs mass quite strongly already at the time at which the LHC experiments started. Based on previous data from LEP (Large Electron Positron Collider) and Tevatron and early LHC data, one could set an upper bound of 141 GeV for the Higgs mass at 95% confidence level before the 127 GeV signal had been measured (ATLAS 2011). In conjunction with the fact that earlier experiments had already directly excluded a Higgs mass of less than 115 GeV at 95% confidence level, there remained a rather small window of plausible Higgs mass values. Therefore, physicists had strong reasons to expect that, if a Higgs particle would be found at all, it should be found between 115 and 141 GeV.

The described theoretical status quo had the potential to influence the evaluation of the empirical data on the Higgs from December 2011. The theoretician's position suggested taking seriously our theoretical knowledge about the Higgs particle. First, it seemed justified to presume on theoretical grounds that a scalar Higgs particle was likely to exist. Second, one knew that if a Higgs particle existed, it was most likely to have a mass between 115 and 141 GeV. A significant excess of Higgs-like events therefore could only count as a serious candidate for a Higgs discovery if it corresponded to a Higgs mass between 115 and 141 GeV. On that basis, however, a signal between 115 and 141 GeV had to be related to a very different LEE than an imagined signal outside that mass window. The latter signal had to be treated as a potential indication of unexpected and not-yet-understood new physics whose appraisal had to account for a strong LEE (let us say, in the given case, 80 bins or even more). An excess rate between 115 and 141 GeV, however, if not a fluctuation, was very likely caused by Higgs particle exchange. Therefore, only that energy range seemed relevant for specifying the LEE in an experiment that was expected to find a Higgs particle. Given a width of the signal of 1 GeV, this corresponded to a factor of 26, which increased the significance of the data from December 2011 by a factor of 3.

Strictly speaking, the argument did not depend on attributing a high absolute probability to the existence of the Higgs particle. It only hinged on the assumption that the existence of the Higgs particle was far more likely than new physics that could account for the observed excess rate but was not a Higgs particle.[9] In the actual Higgs case, however (as arguably in most other conceivable cases where theoretical assessments can give strong indications for a reduction of LEE), the reasoning was based on the cogency of a specific theoretical approach and therefore on attributing a high ab-

9. In a scenario where all new physics other than the Higgs particle were strictly excluded, arbitrarily small prior probabilities of the Higgs model would suffice for reducing LEE to the regime where a Higgs particle could be found.

https://doi.org/10.1086/679179 Published online by Cambridge University Press

solute probability to the implications of that approach. Therefore, our discussion will focus on that type of reasoning.

Contrary to the theoretician's position, the experimentalist's position suggests that we must not take into account any information about the probability of the Higgs mechanism or any other kind of new physics when assessing the data that are collected to confirm it. On that perspective, theoretical 'prejudices' regarding the existence of an object must not be responsible for announcing the discovery of that object. When disregarding all knowledge about the probability of new physics, however, no justification remained for taking the Higgs particle to be more probable than so-far-unknown kinds of new physics that were not constrained to the mass window between 115 and 141 GeV. In other words, there was no justification to reduce the LEE to the Higgs-mass window. The rigid experimentalist's position therefore had to account for the full LEE implied by the experimental setup. It thus did not generate an argument for moving away from the universal 5σ limit.[10]

From the theoretician's perspective, the LEE could be reduced even further based on a more immediate reliance on the assumption of the existence of the Higgs particle. The data collected at the LHC until December 2011 did not merely indicate a signal at 125 GeV. They also excluded with 95% confidence a Higgs mass above 127 GeV. This limit had to be treated differently than the 141 GeV limit discussed above in an important respect. Since the 115–41 GeV mass window was established based on different data than the data that provided the measured signal itself, the fact that the measured signal was found within that window was not a priori guaranteed and, on that basis, could justify a different treatment of the actual data than what a potential measured signal outside that window would have required. It was possible to say that since the signal was within the mass window, it constituted a serious Higgs candidate; therefore, the LEE could be reduced to the energy range compatible with a Higgs observation. In the case of setting an upper mass limit of 127 GeV, this line of reasoning did not work since the data that constituted the signal were also part of the evidence that implied the mass limit. If the overall data had been different and the measured signal had occurred at a different energy, let us say at 135 GeV, those data obviously would not have implied a mass limit of 127 GeV. Therefore, it was not true in the given case that a measured signal above 127 GeV would have been incompatible with a Higgs particle observation. The line of reasoning supporting the reduction of the LEE that

10. Since any specification of LEE in a particular experimental context must rely on some theoretical considerations based on the theory to be tested, the experimentalist's position arguably suggests a general tendency to stick to the default position of a rigid universal discovery condition rather than allow for a conceptual analysis of the individual experimental context.

https://doi.org/10.1086/679179 Published online by Cambridge University Press

was presented in the previous paragraphs thus did not work this time. The LEE could be reduced to the range between 115 and 127 GeV, however, if one was ready to rely on the conviction that the Higgs particle existed with a high probability. Under this assumption, one could adhere to the following line of reasoning:

1. It is likely that the Higgs exists (for the sake of the argument, let us assume the arbitrarily chosen probability of 90%).[11]
2. With a probability of 95%, the Higgs does not have a mass below 115 or above 127 GeV.
3. Therefore, with a probability of $0.90 \times 0.95 = 0.855$ there is a scalar field within that mass window, whether or not any other new physics can be found beyond that window.
4. Assuming that there is a scalar within that mass window, only a signal within that mass window can be identified with it. Therefore, under this condition, it is justified to reduce the LEE to the range between 115 and 127 GeV. Given the high trust in the existence of the Higgs particle (and once again assuming a width of the signal of 1 GeV), we thus can reduce the LEE to a factor close to 12 (with our chosen probability we would have $0.855 \times 12 + 0.145 \times 26 \approx 14$).

The LEE is then reduced by another factor of 2. Since this second step of reducing the LEE is directly based on a high probability of the Higgs hypothesis, it is 'theoretically contaminated' to a higher degree than the first one and thus even less acceptable from an experimentalist's position.

Both steps in conjunction reduced the LEE by a factor of 6. The significance of the data when taking into account the LEE thus would have increased from a less than 3σ effect when seen from the experimentalist's position to an effect well beyond the 3σ limit.[12]

5. The Significance of the Debate. Why is the discussion between the experimentalist's and the theoretician's position of interest? It is not relevant

11. While 90% may seem high, it is not unrealistic. It would not have been difficult in 2010 to find high-energy physicists who would have accepted 9:1 odds for bets on the Higgs discovery at the LHC.

12. Both the full LEE and the theoretically constrained LEE are sometimes accounted for in the experimental literature in stating global *p*-vales. ATLAS (2012, 15), after giving the local *p*-value for the Higgs candidate signal, also states a global *p*-value "taking into account the entire mass range of the search, 110–600 GeV." In CMS (2012, 41), global *p*-values in the search ranges 115–30 GeV and 110–40 GeV are stated. Those are the search ranges determined by constraints from perturbative corrections of other HEP processes and from the Higgs search-based exclusion of Higgs masses, respectively. Neither of the cited papers, however, suggests that global *p*-values could or should be factorized into the 'official' discovery limit that hinges on local *p*-values.

https://doi.org/10.1086/679179 Published online by Cambridge University Press

any more for the assessment of the Higgs data since the evidence collected in 2012 amounted to the discovery of a Higgs particle on any account. In order to understand why the importance of the discussion goes beyond that of an anecdote during the process of Higgs data collection, we first have to define the character of the debate at a slightly more general level. The question at stake is whether it makes sense to specify discovery limits individually for specific experimental searches based on knowledge about and trust in the theory to be tested in the given cases.

This question emerged in the context of the Higgs search for the first time as a serious matter of debate. At early stages of HEP experimental testing, there were general debates on meaningful discovery limits, but they were not addressed from this specific angle. Two reasons are mainly responsible for the fact that the topic became interesting in the context of the Higgs search.

First, there is the issue of time frames. As long as the process of data collection for establishing a new particle is a matter of hours, days, or even a few weeks, it does not make much sense to start a 'philosophical' discussion about the status of an intermediate result of a specific data-collecting process or to analyze the significance of the data by analyzing the details of LEE in the given case. Applying a general rule of thumb, which eventually crystallized into the rigid 5σ discovery criterion, thus used to be the method of choice for everyone involved. The search for the Higgs particle, given its technical and conceptual complexity, required fairly long timescales. The announcement of evidence for the Higgs particle came three years after the experiment had started, and it took another seven months until a discovery could be announced. This was a sufficiently large time frame for justifying an official announcement of evidence for the Higgs particle that did not yet amount to a discovery, which made an individual discussion of statistical aspects of the Higgs case plausible.

The second issue is the degree of trust in the theory to be tested. As discussed in section 4, a (comparably) high degree of trust in a theory's predictions is a precondition for the theoretician's position to make sense. In early HEP, experiments mostly discovered new particles that had not been predicted by theory at all. Only with the establishment of internal symmetries and gauge theory did theory become sufficiently strong for making physicists highly confident about the existence of conjectured particles before they were experimentally discovered. And only on that basis did the distinction between the experimentalist's and the theoretician's position become meaningful.[13]

13. Both stated conditions were fully fulfilled during the search for the top quark. The reasons why the debate on LEE did not arise in that case have to do with the specifics of the experimental setup, which were not favorable to a discussion of LEE.

https://doi.org/10.1086/679179 Published online by Cambridge University Press

Long time frames of empirical searches and considerable levels of trust in empirically unconfirmed theories are characteristic of the overall evolution of fundamental physics today. Supersymmetry (SUSY), which constitutes a possible structural characteristic of HEP that would imply the existence of a wide range of new elementary particles and may be found at the LHC within the next decade, would be far more difficult to establish conclusively than the Higgs particle. It would presumably take several years to get from the first indication of SUSY to an announcement of its conclusive confirmation.[14] Empirical confirmation of other, more far-reaching theoretical hypotheses in fundamental physics such as cosmic inflation or string theory, to the extent it will be achieved at all, must be expected to require even longer time frames and presumably will be based on complex patterns of cosmological data that are less conclusive than the data extracted from collider physics.

In all those cases, theoretical arguments make theoreticians instill a certain level of confidence in the viability of the respective theories (though none of the mentioned cases make physicists as confident as they were in the Higgs case). Fundamental physics thus has entered a stage where theories are considered likely viable first for theoretical reasons, and if any evidence for them is going to be discovered, the presence of significant but inconclusive evidential support for individual theories will constitute the status quo for many years or even decades. The discussion about the status of the Higgs data in this light may be seen as a test case for a debate that is due to arise with more urgency in the foreseeable future: to what extent is it legitimate to move away from the canonical experimentalist's position in cases where physics (1) can rely on a strong and cogent theoretical analysis of the overall physical context and (2) must specify the status of its theories based on inconclusive data for periods of time that may approach the length of an individual scientific career? To be sure, the specifics of the discussion of the Higgs discovery case will not fully apply to most new contexts in HEP and cosmology. Nevertheless, the Higgs case constitutes a first example where a specific way of going beyond the experimentalist's position can be laid out. It is on that basis that the debate on the Higgs evidence merits closer philosophical inspection.

6. Bayesianism versus Frequentism. The debate between the experimentalist's and the theoretician's position can be understood in terms of a con-

14. This long time frame is caused by the complex problem of identifying specific particles as an expression of SUSY once they are discovered. A potential modified assessment of LEE with respect to the discovery of an individual SUSY candidate would not significantly reduce that time frame. Still, if SUSY candidates are discovered at all, situations reminiscent of the Higgs case may emerge at later stages of the search for SUSY particles.

https://doi.org/10.1086/679179 Published online by Cambridge University Press

flict between a rigidly frequentist and a partially Bayesian perspective. Let us first briefly characterize a Bayesian and a frequentist perspective in their pure forms. The Bayesian approach aims at extracting probabilities of the truth of a scientific hypothesis H in the face of empirical evidence E based on the probabilistic relation

$$P(T|E) = P(E|T)\frac{P(T)}{P(E)},$$

where T denotes the statement that hypothesis H is true. Empirical evidence E is taken to confirm H iff it increases the probability that H is true, that is, iff

$$P(T|E) > P(T).$$

The prior probability $P(T)$ can be informed by old empirical data, theoretical considerations, or prejudice. The posterior extracted from considering one set of data E can serve as a prior for the next step of empirical testing. To get started with the empirical process, however, the scientist has to assume a prior that is not based on empirical tests of the given hypothesis. For that reason, the Bayesian perspective always contains a subjective element.

From a Bayesian perspective, the scientific process can nevertheless be seen as intersubjectively reliable. This conclusion is based on a specific property of Bayesian theory confirmation: under plausible conditions, repeated consideration of new empirical data leads to converging posterior probabilities. Starting from very different prior probabilities for a theory H that offers statistical predictions of E, after a sufficiently extensive series of empirical tests, agents end up with very similar posteriors if they agree on the empirical implications of the theory's alternatives H_i.[15]

The frequentist approach has a different focus than Bayesianism. While Bayesianism aims at modeling a full and coherent argumentative structure that starts from the assessment of the probability of prior assumptions and ends with the assessment of the probability of a hypothesis in the face of all known data, the frequentist approach aims at isolating the statistical analysis of numerically well specified empirical data from vague assessments of priors. The frequentist carries out a statistical analysis of the empirical data within a given conceptual framework that is taken for granted but without relying on any information about the prior probability of the hypothesis to be tested. On that basis, precise statistical analysis can, within a given conceptual framework, determine a theory's likelihood, which directly corresponds to P

15. This includes the so-called catch-all hypothesis.

https://doi.org/10.1086/679179 Published online by Cambridge University Press

$(E|T)$.[16] From the likelihood the theory's p-value can be extracted, which in turn is taken as an indicator of the theory's viability.

The problem of the frequentist approach is that it does not specify the probability of the tested hypothesis. There is an ongoing debate on whether—and if so, how—frequentist analysis nevertheless supports an inductive inference to the tested hypothesis. One influential philosophical position, developed by Deborah Mayo and Aris Spanos (Mayo 1996; Mayo and Spanos 2006) and applied to HEP by Kent Staley (Staley 2004), suggests that a frequentist analysis allows for the assessment of the severity of an empirical test. On that account, a hypothesis that has survived severe testing can be inductively inferred without any reference to the probability of the hypothesis. Most experimental high-energy physicists, however, implicitly assume an epistemic Bayesian embedding of their frequentist analysis. That is, they think that a signal that is highly significant based on frequentist data analysis can be interpreted in terms of a high probability of the viability of a corresponding physical hypothesis. This can be justified to a certain extent by the convergence behavior of $P(T|E)$ under repeated empirical testing. On this view, the experimentalist who infers a hypothesis based on frequentist data analysis relies on the understanding that the impact of subjective priors is eventually 'washed out' by empirical data.

Applied to the Higgs search at the LHC experiments, the frequentist analysis takes for granted the empirically well confirmed parts of the standard model of particle physics plus a wide range of scientific concepts in collider physics as the basis for analyzing the data. It does not, however, admit any information that relies on the assessment that the Higgs particle is likely to exist. The frequentist perspective therefore resembles the canonical experimentalist's position.

A Bayesian perspective on the Higgs search, to the contrary, would acknowledge that the analysis of specific experiments at the LHC may be based on informed priors for the existence of the Higgs particle. These priors are themselves based on previously collected empirical data and theoretical reasoning relying on those data. Since a long and elaborate scientific process has led up to the emergence of those priors, they must not be understood as entirely subjective, even though the specific values attributed to them by individual scientists may differ to some degree as a result of subjectively chosen presumptions. In other words, the scientists' trust in the existence of the Higgs particle before the start of LHC experiments can be understood as the result of a scientifically informed assessment rather than of mere subjective prejudice.

16. Unlike the Bayesian $P(E|T)$, the frequentist likelihood is no conditional probability. To mark that difference, it is normally written in a slightly different notation than the Bayesian $P(E|T)$. The distinction is of no importance for the discussion of this paper, however.

https://doi.org/10.1086/679179 Published online by Cambridge University Press

A full realization of a Bayesian perspective on the Higgs search would need to extract the posterior probability of the null-hypothesis $P(T_N|E)$ from $P(E|T_N)$ by specifying the prior probability $P(T_N)$ and $P(E)$. In order to give a quantitative example, let us imagine that physicists collect empirical evidence at the LHC that implies $P(E|T_N) = 0.01$. The Bayesian now takes into account that we strongly believed in the Higgs particle already before the experiment and, let us say, attributed a probability of 90% to its existence. Therefore, we excluded the null hypothesis with at least 90% probability from the start. Assuming the simplified scenario that the Higgs hypothesis is the only plausible alternative to the null hypothesis, we write $P(T_N) \approx 0.1$.[17] Given that we strongly expect E if the Higgs exists, we have $P(E) \approx 0.9$, which gives

$$P(T_N|E) = P(E|T_N)P(T_N)/P(E) \approx 0.01 \times (1/9) \approx 0.0011.$$

(For the sake of simplicity, we have ignored LEE at this point. It will enter the picture below.) A Bayesian approach thus would generate a considerably higher degree of trust in the Higgs particle than the frequentist statistical analysis of the numerical data.

Though Bayesian approaches are being discussed in experimental HEP in order to get a better understanding of the overall process (see, e.g., Cousins 1995, 2013; Read 2002; Lyons 2013a), no physicist would propose to replace the frequentist statistical analysis by a consistently Bayesian line of reasoning that fully accounts for the physicists' prior expectations. The motives for that restraint are very clear. By establishing the Bayesian analysis of the research process as a fully viable strategy of scientific data analysis, one would permit that a rigid quantitative statistical analysis where the numerical input is well determined by the empirical data gets adulterated by probability assessments that are vague and subjective. It would put guessing priors on the same footing as rigid and quantitative experimental testing.

The theoretician's perspective constitutes a less intrusive way of introducing aspects of Bayesian reasoning into data analysis. The approach remains within the general framework of frequentist testing. Therefore, it avoids messing with the objective character of statistical data analysis itself. However, it acknowledges the general relevance of attributing probabilities to the truth or viability of theories and therefore, in principle, accepts the viability of Bayesian reasoning. On that basis, a defender of the theoretician's perspective will presumably endorse the idea that frequentist hypothesis testing can justify inference to the tested hypothesis based on the Bayesian convergence theorems (see above). Being sympathetic to

17. Attributing a significant probability to other kinds of new physics while keeping the trust in the Higgs at 90% would further decrease the posterior probability of T_N.

https://doi.org/10.1086/679179 Published online by Cambridge University Press

Bayesian reasoning while remaining sceptical about allowing it to corrupt the objective character of data analysis naturally suggests using elements of Bayesian reasoning when thinking about discovery criteria, which, as emphasized in section 3, are based on subjective choices anyway.

Let us briefly restate the theoretician's perspective on LEE in a Bayesian framework. The discovery criterion specifies a certain significance level of the data, which corresponds to a certain local p-value. This local p-value is calculated by integrating the probabilities $P(E|T_N)$ for excess rates E at least as high as the measured rate. The choice of a meaningful discovery criterion, however, must depend on the global p-value, which takes into account LEE. The global p-value is extracted from the local p-value by multiplication with a look elsewhere factor $\langle \mathrm{LEE} \rangle$ that, in the given case, corresponds to the number of energy bins where a signal could have been found just as well. Now the theoretician's position consists in assessing the size of LEE based on the prior expectation that the Higgs particle probably exists. By specifying prior probabilities for the truth of the Higgs hypothesis, we can thus explicitly weight the reduced LEE by our degree of trust in the Higgs hypothesis. The most straightforward way of doing this is to write[18]

$$\langle \mathrm{LEE} \rangle = [\langle \mathrm{LEE} \rangle_{T_H} P(T_H) + \langle \mathrm{LEE} \rangle_{T_{\sim \mathrm{H}}} P(T_{\sim \mathrm{H}})]. \quad (1)$$

$\langle \mathrm{LEE} \rangle_{T_H}$ and $\langle \mathrm{LEE} \rangle_{T_{\sim H}}$ denote the LEE factors under the assumption that the Higgs hypothesis is true and under the assumption that it is false, respectively. $P(T_{\sim H}) \geq P(T_N)$ must be satisfied. Strong trust in the truth of hypothesis H (i.e., high $P(T_H)$) in conjunction with a factor $\langle \mathrm{LEE} \rangle_{T_H}$ that is reduced compared to the factor $\langle \mathrm{LEE} \rangle_{T_{\sim H}}$ implies a reduced value of $\langle \mathrm{LEE} \rangle$ and therefore a reduced global p-value.

The theoretician's perspective thus amounts to a partial inclusion of Bayesian reasoning by introducing $P(T_H)$ into the assessment of the data. No Bayesian updating that leads from $P(T_H)$ to a determination of $P(T_H|E)$ takes place, however. Bayesian priors are used solely for assessing the size of LEE and, on that basis, eventually for influencing the assessment of the significance of the collected data. If this strategy were consensually accepted in the research field, it would suggest the quantitative specification of individual discovery limits in specific experimental contexts. Those limits would explicitly rely on global p-values extracted based on the theory-based assessment of the given experimental context. It is a drawback of the described approach that it constitutes a peculiar hybrid of Bayesian and

18. This simple model is not based on a fully Bayesian reconstruction of LEE (which is not carried out in this paper). A fully Bayesian reconstruction would show a stronger dominance of reduced LEE, which would extend even to cases where $P(H)$ is similar in size to $P(\sim H)$ (Dawid 2014).

https://doi.org/10.1086/679179 Published online by Cambridge University Press

frequentist strategies of reasoning. Its main advantage consists in accounting for the Bayesian element of human reasoning to a certain degree without compromising the methodological purity of scientific data analysis in its core regime. Vague and subjective prior probabilities are used only within the interpretational part of data analysis, which is vague and subjective anyway.

In the debate on the December 2011 data, considerations that in effect amounted to the described line of reasoning were made in a fairly informal way in order to justify one's own subjective trust in the data. No one proposed using the described 'weak' deployment of Bayesianism 'officially' for the specification of an LEE-dependent discovery criterion that could replace the rigid 5σ criterion. In principle, however, an altered context of experimentation, where the conceptual background is deemed very trustworthy and the timescales for the collection of conclusive empirical data are very high, might suggest the latter step as a reasonable way to go.

A closer look at the characteristics of empirical testing in HEP further strengthens the case for the deployment of elements of Bayesian reasoning in scientific data analysis. As it turns out, in specific contexts a dogmatic adherence to frequentist principles fails to be consistent with elementary scientific intuitions about theory confirmation. In the following section, an example of this kind shall be discussed.

7. An Argument for Adopting the "Theoretician's Perspective." It was argued above that the plausibility of the 5σ limit directly depends on the LEE that happens to be relevant in today's high-energy experiments. If a significantly higher LEE became typical for experimentation in HEP, statistical fluctuations of 5σ significance would start showing up in experimentation. It may be expected that scientists would then be led to introduce a higher limit for the discovery of a particle in order to maintain the trustworthiness of experimental results. Note that this implication does not depend on whether experimentalists break up a new generation of experimental tests into a very large number of individual experiments that altogether test very large mass ranges or whether they build individual experiments that test very large mass ranges. The crucial point is that the testing of very large mass ranges increases the probability of measuring very large fluctuations somewhere in the measured ranges. To give a specific example, if a new generation of experiments looked for new particles in 10,000 bins, the chances of finding a 5σ fluctuation in one specific large experiment of that kind would be about 3×10^{-3}. It would clearly be premature in such a context to call a 5σ signal a discovery of a corresponding particle. A 6σ limit would be necessary to retain the old level of trustworthiness of experimentation in the new scientific context.

Let us now imagine a theory H_2 that predicts the existence of a scalar particle h with a mass within a parameter interval I_1 and forbids the exis-

https://doi.org/10.1086/679179 Published online by Cambridge University Press

tence of a scalar particle with a mass within the 100 times larger interval I_2. In order to test H_2, physicists first build an experiment that scans interval I_1 with 100 bins. They succeed in finding a scalar with 5σ confidence level, which, as experimentation still proceeds within the 'old' context of testing smaller energy ranges and therefore of a smaller LEE, constitutes the criterion for calling the data a discovery at the time. Later on, experimentalists develop new techniques that allow them to test far wider energy ranges. On that basis, they build a larger series of experiments in order to test the second prediction of H_2. In those experiments, they scan the interval I_2 with 10,000 bins (without providing further tests of I_1) and find no further scalar, which is in perfect agreement with H_2. Now physicists proceed to carry out an overall analysis of the entire set of data collected in all experiments. Since experimentation has entered a stage of testing wider energy ranges, it would be necessary to correct the limit for discovery and shift it, say, to 6σ. The first experiment, however, does not differ structurally from the later experiments. Therefore, it would seem awkward to apply a different discovery criterion there than in the other experiments just on the basis that it came first. Once the new criterion has been established, it would seem plausible to apply the new 6σ criterion to the old experiment as well. However, the data that indicate a scalar field in interval I_1 do not amount to a 6σ effect, which means that physicists then would not be allowed to speak of a discovery of the scalar h anymore. The confirmation of theory H_2 by the observation that no scalar exists in I_2 thus, in effect, would have invalidated the discovery of the phenomenon h that is predicted by H_2. An implication of this kind clearly is at variance with our intuitive understanding of confirmation and discovery.

The theoretician's perspective avoids the above implication because the trust in theory H_2 blocks the LEE with respect to I_2. The theoretician's perspective implies the following line of reasoning: the experimental testing of I_1 has led to the discovery of scalar h and thereby has established the viability of H_2 with high probability; since H_2 predicts that no scalar particles within the mass range I_2 exist, strong trust in H_2 implies that I_2 has little relevance for LEE with respect to the search for scalar particles; the experimental testing of I_2 therefore does not significantly change the discovery condition for h and cannot invalidate the discovery of h. By avoiding the paradoxical conclusion suggested by the experimentalist's position, the theoretician's position looks decidedly more plausible than the former in the given context.[19]

19. One may speculate that an experimentalist facing the described situation would indeed implicitly adopt the theoretician's perspective and, on that basis, keep the 5σ criterion with respect to the first experiment.

https://doi.org/10.1086/679179 Published online by Cambridge University Press

REFERENCES

ATLAS. 2011. The ATLAS and CMS Collaborations. ATLAS-CONF-2011-157.
ATLAS Collaboration. 2012. "Observation of a New Particle in the Search for the Standard Model Higgs Boson with the ATLAS Detector at the LHC." *Physics Letters* B 716:1–29.
CMS Collaboration. 2012. "Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC." *Physics Letters* B 716:30–61.
Cousins, Robert D. 1995. "Why Isn't Every Physicist a Bayesian?" *American Journal of Physics* 63:398.
———. 2007. "Annotated Bibliography of Some Papers on Combining Significances or p-Values." PhilSci Archive, Cornell University. arXiv:07052209v2.
———. 2013. "The Jeffrey's-Lindley Paradox and Discovery Criteria in High Energy Physics." *Synthese*, online first, doi:10.1007s11229-014-0525-z.
Dawid, Richard. 2014. "Bayesian Perspectives on the Discovery of the Higgs Particle." Paper presented at the "Evidence, Discovery, Proof: Measuring the Higgs Particle" workshop, University of South Carolina, April.
Englert, Francois, and Robert Brout 1964. "Broken Symmetry and the Mass of Gauge Vector Mesons." *Physical Letters Review* 13:321–23.
Feldman, Gary J. 2006. "Concluding Talk: Physics." Paper presented at PHYSTAT 05, "Statistical Problems in Particle Physics, Astrophysics and Cosmology," Oxford, September 12–15, 2005.
Franklin, Alan. 2013. *Shifting Standards: Experiments in Particle Physics in the Twentieth Century.* Pittsburgh: University of Pittsburgh Press.
Higgs, Peter. 1964. "Broken Symmetries, Massless Particles and Gauge Fields." *Physics Letters* 12:132–33.
Karaca, Koray. 2013. "The Construction of the Higgs Mechanism and the Emergence of the Electroweak Theory." *Studies in History and Philosophy of Science* B 44:1–16.
Lyons, Louis. 2013a. "Bayes and Frequentism." PhilSci Archive, Cornell University. arXiv:1301.1273.
———. 2013b. "Discovering the Significance of 5 sigma." PhilSci Archive, Cornell University. arXiv:1310.1284.
Mayo, Deborah. 1996. *Error and the Growth of Experimental Knowledge*. Chicago: University of Chicago Press.
Mayo, Deborah, and Aris Spanos. 2006. "Severe Testing as a Basic Concept in a Neyman-Pearson Philosophy of Induction." *British Journal for the Philosophy of Science* 57:323–57.
Read, A. L. 2002. "Presentation of Search Results: The CLS Technique." *Journal of Physics* G 28:2693–704.
Staley, Kent. 2004. *The Evidence for the Top Quark: Objectivity and Bias in Collaborative Experimentation*. Cambridge: Cambridge University Press.

https://doi.org/10.1086/679179 Published online by Cambridge University Press