CrossMark

Bayesian perspectives on the discovery of the Higgs particle

Richard Dawid[1]

Abstract It is argued that the high degree of trust in the Higgs particle before its discovery raises the question of a Bayesian perspective on data analysis in high energy physics in an interesting way that differs from other suggestions regarding the deployment of Bayesian strategies in the field.

Keywords High energy physics · Higgs particle · Discovery · Confirmation · Bayesianism · Frequentism · Non-empirical evidence

1 Introduction

Physicists had a remarkably high degree of trust in the existence of the Higgs particle long before it was discovered at the LHC.[1] It shall be argued in this paper that this specific phenomenon opens up interesting perspectives on the role of Bayesian statistics in experimental high energy physics (HEP).[2]

Data analysis can be carried out within two different conceptual frameworks. Frequentist data analysis provides likelihoods $\mathcal{L}(H|E)$, and p values[3] of a hypothesis H in dependence on data E. The likelihood $\mathcal{L}(H|E)$ is equal to the probability $P(E|H)$

---

[1] This denotes the idea that some kind of Higgs-like particle existed. Whether the Higgs sector was standard-model-like, supersymmetric, composite or else remained an open question.

[2] To be sure, the Higgs particle is not the first example of a prediction in HEP that was strongly believed before empirical confirmation. Still, it offers a particularly nice example for discussing the probabilistic status of a hypothesis prior to empirical testing.

[3] We will come back to p values in Sect. 3.

✉ Richard Dawid
richard.dawid@univie.ac.at

[1] Center for Mathematical Philosophy, Ludwig Maximilian University Munich, Geschwister Scholl Platz 1, Munich, Germany

of data E given the truth (or viability) of hypothesis H. Bayesian data analysis, to the contrary, provides probabilities $P(H|E)$ of hypothesis H given data E. Unlike frequentist analysis, Bayesianism thus offers a direct assessment of the chances that a hypothesis H be true or viable.

The canonical approach towards empirical testing in HEP is based on frequentism. The advantages of frequentism from a working physicist's perspective are clear. Frequentism provides a framework for carrying out quantitative data analysis without considering subjective prior probabilities. Likelihoods and p values of hypothesis H for data E can be deduced from the hypothesis itself, which allows extracting quantitative results in a univocal way from measured data within a well-defined framework of theory-based assumptions. To be sure, this does not mean that the *decision* whether to endorse or reject a hypothesis is free from subjective considerations. That decision is based on setting a significance limit for discovery. Choosing that limit is based on subjective assessments of various kinds. However, those subjective assessments do not interfere with the quantitative data analysis itself.

Bayesian data analysis extracts $P(H|E)$ from $P(E|H)$ based on the Bayes formula

$$P(H|E) = P(E|H)\frac{P(H)}{P(E)} \qquad (1)$$

This approach relies on subjective prior probabilities $P(H)$ and $P(E)$, which introduce an element of subjective choice into data analysis that is absent in the frequentist case. Generally speaking, there can be two kinds of motivation for looking at Bayesian data analysis despite that disadvantage. First, one may find frequentist data analysis epistemologically unsatisfactory since it does not specify a probability for a theory's truth or viability. At a more pragmatic level, one finds contexts where a straightforward application of frequentist data analysis fails to provide the best characterisation of the impact of some data E. Bayesian methodology can then be deployed in a target-oriented way in order to optimize the results of data analysis.

Ideas to make use of the Bayesian perspective in HEP up to this point have been of the second kind. Bayesian methods have been used, as we will see more specifically in the next section, on a purely pragmatic basis while keeping up the frequentist character of overall data analysis. The present paper makes the point that the case of the conjecture and eventual discovery of the Higgs particle gives reasons for going beyond such a purely pragmatic deployment of Bayesian methods and think about an overall Bayesian embedding of HEP data analysis. This does not imply that HEP data analysis should switch to Bayesian methodology. It does mean, however, that keeping in mind a Bayesian framework can help making the right choices in framing data analysis.

The case of the Higgs search is of particular interest for our analysis because of the high degree of trust in the Higgs hypothesis long before the Higss particle's empirical discovery. As long as one is testing theories whose viability remains unclear at the time of empirical testing, disregarding prior probabilities and having a frequentist perspective on endorsing or rejecting the theory based on the incoming data does not encounter the problems we are going to discuss. In cases where a hypothesis is either

taken to be very probably viable or most probably false already before empirical testing, however, disregarding those convictions may lead to an inadequate understanding of the results of data analysis. Attributing a substantial prior probability to the hypothesis constitutes a natural way of dealing with situations of this kind. The high level of trust in the Higgs particle before it was discovered makes it a prime candidate for applying that strategy.[4]

In order to contrast the suggested role of a Bayesian perspective on data analysis in the Higgs case with the way Bayesian methods have been used in HEP before, we start with a brief look at an example of the latter kind in Sect. 2. Afterwards, Sects. 3, 4 and 5 analyse the Higgs case. Section 6 finally compares the two contexts and makes an attempt to specify the significance of Bayesian reasoning in the Higgs context (and beyond).

Before entering the specific analysis, however, it seems adviseable to say a few more words about the epistemic relation between frequentist and Bayesian data analysis.

It is clear that a "philosophical" justification for assessing a theory's viability based on frequentist data analysis must rely on *some* wider epistemological framework: it has to be clarified on what basis the likelihoods, p values and error probabilities that are calculated in a frequentist framework are supposed to justify the endorsement or rejection of a given hypothesis. The debate on how this should be best understood goes back to the calssical works of Fisher, Neyman and Pearson. In recent philosophy of science (Mayo 1996; Mayo and Spanos 2006; Staley 2004) have argued that the endorsement or rejection of theories can and should be justified based on the concept of the "severity of testing" without any reference to the probability of a theory's truth. Others have argued that a full understanding of the epistemic status of frequentist data analysis can only be acquired by relating a theory's likelihood to its probability, which amounts to embedding frequentist data analysis within a framework of Bayesian epistemology. (see e.g. Howson 2000.[5])

It is probably fair to say that most statisticians in HEP (and other research fields) think that frequentist data analysis does require a Bayesian epistemic embedding. Nevertheless, they normally refrain from making that embedding explicit within data analysis itself. There is a wide consensus in the field that data analysis should be carried out in a way that makes the published quantitative results reproducible by anyone who

---

[4] Another excellent recent example is the case of a hypothesis that was considered most likely false despite seemingly significant empirical data in its support. When the OPERA experiment seemed to measure superluminal neutrinos in 2011, the very low prior probability of the hypothesis that neutrinos could be faster than light (due to very strong, data-based and conceptual trust in the viability of the theory of special relativity) led to a situation where conventional frequentist methods of data assessments were disregarded by most physicists in their assessment of the OPERA data and an implicitly Bayesian assessment of the situation took over. While the prior probability of a mistake in the OPERA experiment was arguably fairly low, it was still considered much higher than the prior probability of neutrinos moving faster than light. In effect, Bayesian updating thus led most physicists to infer from the data an experimental mistake rather than the truth of the hypothesis.

[5] Howson points out that Mayo's strategy achieves decoupling from the theory's truth probability only at the price of becoming insensitive to the threat of the so-called base-rate fallacy. Therefore, Howson argues, Mayo's reasoning cannot justify trust in a theory that has been "severely tested" in Mayo's sense.

knows experiment, data and the theory that is tested.[6] The introduction of priors puts this inter-subjective quality of the published data into jeopardy. In some special cases, however, HEP-statisticians have suggested Bayesian elements of reasoning in order to extract satisfactory quantitative results from the data. Examples are Robertson and Knapp (1988), Cousins (1995), Read (2002) and Lyons (2013). Clearly, Bayesian ideas were deployed in those cases with the principle of a general Bayesian embedding of data analysis in mind. The explicit goal, though, was to use Bayesian strategies in a limited way that kept subjective assessments of prior probabilities out of the game as much as possible. In the following section, we have a look at one particular example of this strategy.

2 Excluding unphysical parameter values

It sometimes happens that a data-based frequentist determination of the expected value of a given parameter suggests a value that is physically meaningless or impossible. A nice example was the measurement of neutrino masses in the 1980s based on tritium beta decay - well before the eventual conclusive measurement of non-zero neutrino masses. The tritium beta decay rate had been calculated to depend on the squared mass of the electron neutrino generated in the decay process. The probability of emitting an electron with total energy E in beta decay was found to be

$$dN(E) = K|M|^2 F(Z, R, E) p_e E(E_0 - E)((E_0 - E)^2 - m_\nu^2 c^4)^{1/2} dE \qquad (2)$$

with a term K containing some coupling constants, a transition matrix element M, a Coulomb correction term F, the electron momentum p_e and the total energy E. Of interest for our purposes is solely the dependence on the squared neutrino mass m_ν^2. Based on fundamental laws of physics, a neutrino mass squared term only made physical sense if it was larger than or equal to zero. Due to statistical fluctuations, however, the measured decay rate could lead to likelihoods $\mathcal{L}(m_\nu^2|E)$ which had the highest values for negative m_ν^2. In the given case, the experimental results in the late 1980s gave a confidence interval of $m_\nu^2 = (-54 \pm 30)eV^2$. Statisticians thus faced the question how to deal with a result that put the entire 1σ confidence interval in the physically meaningless regime. Apart from the counter-intuitive appearance of that result, the problem arose how to extract the optimal physical interpretation from the data. The issue of primary interest addressed by the experiment was to specify an upper bound on the electron neutrino mass. Even when choosing a significance level that was sufficiently high for putting the upper bound for m_ν^2 above zero, that bound seemed misleading for the following reason: one knew from the fact that the likelihoods peaked well below zero that the data was statistically off towards too low m_ν^2 values and thus led to an underestimation of the value of the upper mass-square bound. But a straightforward frequentist analysis provided no way to account for that knowledge.

---

[6] It will be important for the analysis of later sections that this requirement can't be upheld for the inter-pretation of the published data.

Several possible remedies for this problem were discussed. One of those suggestions, realized in Robertson and Knapp (1988) and discussed e.g. in Cousins (1995), adopted a Bayesian perspective in order to limit the statistical analysis to the physically allowed parameter region. This was done by attributing prior probability zero to the physically meaningless negative mass square values and assuming a uniform distribution of prior probabilities for physically meaningful mass-square values. On that basis, one could extract from the same data as before a positive expectation value for the squared neutrino mass and a higher bound on the neutrino mass.

Two conceptual issues come along with the suggested use of Bayesian methodology, however. The first one is of immediate relevance for extracting probabilities of physical hypotheses from the data and was discussed in Cousins (1995). As mentioned above, the Bayesian analysis in the given case relies on assuming a uniform prior probability distribution for the allowed parameter region. It is by no means clear, however, with respect to which entity the distribution should be uniform. Coming from the formula for beta decay, the most natural choice would be a uniform probability distribution over the parameter m_ν^2. But one might also assume a uniform probability distribution over m_ν based on the argument that m_ν is the basic physical concept in the given case. One might even find reasons for making other choices that are different from both previous ones. Each choice leads to a different upper bound for the neutrino mass. Conceptually, what lies behind this problem is the fact that we have no cogent basis for attributing specific probabilities to specific mass values. Our epistemic position might just as well be characterized as a *suspension of judgement* with respect to the neutrino mass within the physically allowed region. A suspension of judgement, however, cannot be expressed in terms of probability distributions in a Bayesian framework. (See e.g Norton 2008 for a discussion of that point.)

The second point is of a more general epistemic nature. The use of Bayesian methodology along the described lines is a helpful and viable methodological move in data analysis. Statisticians had good reasons for trusting Formula (2) and for working within its framework. On that basis, it made sense to apply Bayesian methods in order to account for the impossibility of negative mass squares.

It is important to emphasise, however, that the way Bayesian methodology is deployed in the given case does not reflect a genuinely Bayesian *epistemic* perspective on data analysis. For the (canonical) Bayesian epistemologist, the interesting and genuinely scientific part of the process of probability assessment is the updating of probabilities under incoming data. Prior probabilities are necessary for getting the process started but should never dogmatically predetermine the outcome of the process. In the given case, however, that is exactly what is happening. By setting to zero the probabilities of those true decay rates which correspond to negative mass square values according to Formula (2), one disregards the possibility that Formula (2) could be incorrect in a way that dissolves the rigid connection between the given true decay rates and negative mass square terms. In other words, setting the probabilities of those true decay rates to zero is based on implicitly assuming probability 1 for the viability of Formula (2). A fully Bayesian epistemic approach to the scenario,

however, would avoid attributing probability 1 to any theory relevant to the data.[7] That would correspond to attributing a probability slightly lower than 1 to Formula (2), which in turn would mean that measured decay rates which correspond to negative mass square values according to Formula (2) would lower the probability of the viability of Formula (2).[8] If the trust in Formula (2) is very high, priors for true decay rates corresponding to negative mass squares based on Formula (2) would still be very low. As long as the significance of the incompatibility between measured decay rates and Formula (2) were not very high, the resulting a posteriori probabilities for the "positive mass square" region would be very similar to the results given by deploying the "zero probability" method. The important point, however, is that the method chosen by the physicists was based on avoiding a non-generic assessment of the actual degree of trust they had in the viability of Formula (2) in the given context. Thereby, they avoided a fully Bayesian epistemic perspective on the physical context they analysed.

Applying Bayesian methods in order to exclude negative mass squares thus has a peculiar tinge. It consists in choosing one probability value (zero) that would not be used in an epistemically Bayesian spirit at all and specifying another one (uniform distribution based on a selected measure) whose specification is not justified by our knowledge of the system. It thus amounts to the use of prior probabilities precisely in those contexts where they seem to have no satisfactory epistemologically Bayesian interpretation. This does not discredit the given strategy as a viable method of data analysis. But Bayesian reasoning as deployed within this framework must be understood as a mere technical tool and does not amount to a coherent epistemically Bayesian viewpoint on data analysis. The overall mindset in the given case remained frequentist and was just "enriched" by Bayesian elements in order to extract optimized quantitative results.

The analysis discussed in this section is an example of the 'objectified' approach to Bayesian data analysis normally adhered to by Bayesians in HEP. This approach avoids addressing the individual subjective element of prior probabilities by relying on generic choices for priors that can be used universally in given types of analysis: a uniform probability distribution for allowed parameter regions, probability zero for excluded parameter regions, probability 1/2 for a null hypothesis, or the like.

In the following sections, we will argue that the significance of Bayesian reasoning in the context of Higgs physics is of a different kind. An explicit quantitative use of Bayesian analysis may actually be less advantageous in the Higgs context than in the case that was described in the present section. However, in order to obtain a

---

[7] It would mean going too far to assert that probabilities zero or "zero for all practical purposes" must not occur in Bayesian analysis. If a hypothesis H is directly inconsistent with a certain empirical outcome E, Bayesian updating under data E sends the posterior probability $P(H|E)$ to zero. We may call that "zero for all practical purposes" because of the Duhem-Quine thesis, which rejects the idea of an inconsistency between data and an individual hypothesis. One may always save the hypothesis by rejecting other, albeit well entrenched, hypotheses. But one may in certain cases deem that possibility negligable for all practical purposes. The step of setting to zero negative mass square values is of a more problematic kind, however, since it sets to zero the probability of an experimental outcome, namely of certain electron emission rates.

[8] Note that this form of Bayesian analysis would indeed correspond to actual reasoning if measured "negative mass square" values became overly significant. In that case, physicists would start wondering whether something is wrong with Formula (2) or the basic physical laws that forbid negative mass squares.

conceptually coherent understanding of data analysis it seems helpful to choose an epistemically Bayesian perspective that is based explicitly on an individually chosen subjective prior that expresses the high degree of trust in the Higgs hypothesis before the discovery of the Higgs particle.

3 Assessing the look elsewhere effect

In July 2012, CERN announced the discovery of a Higgs-like scalar particle (ATLAS 2012; CMS 2012). Subsequent closer data analysis (somewhat) conclusively identified this particle as a Higgs particle. An earlier announcement in December 2011 had already stated an overall signal of close to 4 σ combined significance. The well established convention in HEP to call a signal a discovery only once it exceeds 5 σ significance implied that the December 2011 data was not called a discovery but only "significant evidence" for a Higgs-like particle in official announcements. "Unofficially", however, the status quo after December 2011 gave rise to a debate on the actual significance of the data. This debate was analysed from a philosophical perspective in Dawid (2015). In the following paragraphs, we will rehearse that analysis. Later on, we will embed it in a fully Bayesian framework.

Roughly, one could distinguish two positions in the 2011 debate. On the one side, there were those who in effect stuck to the letter of the discovery criterion and warned of exaggerated trust in data that was insufficient according to that criterion. We will call that position the *experimentalist's position*. On the other side, there were those who argued that the fact that the observed data fit in so nicely with the existing knowledge from theoretical analysis and older data made that present data substantially more trustworthy than it would seem at first sight (more trustworthy, that is, than those 4-σ signals which had turned out to be statistical fluctuations in other contexts in the past.) This position shall be called the *theoretician's position*.

It was argued in Dawid 2014 that the described difference of opinion could in effect be understood as a controversy between the adherents to a rigid frequentist approach to data analysis (who endorsed the *experimentalist's position*) and those who admitted elements of Bayesian reasoning.

In order to understand this point, we first have to look at some basics of the Higgs search and the role of the look elsewhere effect (LEE) in data analysis. The LHC experiments count signatures which could have been created by a new scalar particle. The number of those candidate events is then compared to the calculated background, that is to the expected number of signatures which look the same on a scattering picture but are not generated by an exchange of the new particle that is searched for. Since all involved processes are quantum processes, the numbers of background events are probabilistically distributed. The farther the observed number of events lies above the expected background, the less likely it is that it consists of background events alone.

The first step in HEP data analysis is the test of a null hypothesis N that asserts that no so far unobserved particles contribute to the generation of the observed data E. Data analysis is carried out in a frequentist framework in terms of p values. The local p value of a measured signal (that is of a certain number of signatures of a given type) with a specific characteristic parameter value (corresponding to a certain mass

of the scalar particle) expresses the probability that a signal of at least the measured significance is produced when the null hypothesis is true.

An instructive way of understanding local p values is in terms of error probabilities. The statement: "The measured signal S has p value p" is equivalent to the statement: "assuming that the null hypothesis is true, the strategy of rejecting the null hypothesis based on any signal as strong or stronger than S has a probability p of mistakenly rejecting the null hypothesis" (that is of making a type I error).

The local p value does not constitute a good basis for assessing the genuine significance of a signal, however. Data analysis should reject a null-hypothesis only if the probability of making a type I error (the error of rejecting a hypothesis that is true) in the entire experimental run is very small. The nature of HEP experiments, however, make local p values the wrong number for characterizing that probability. Experiments like ATLAS or CMS at the LHC search for new particles within a wide range of possible mass values. Anywhere within the tested parameter range, statistical fluctuations of the size of the observed signal could arise. So the probability that a true null hypothesis is rejected based on a given significance criterion in a full experiment is related to the probability of the occurrence of a fluctuation of the given strength *anywhere* in the entire tested parameter range rather than at one specific parameter value. This effect is called the look elsewhere effect (LEE).

Obviously, the probability to find a fluctuation anywhere within the tested parameter range is much higher than the probability to find a fluctuation of that size specifically at the energy scale where the signal was actually found. Roughly, the factor that relates the two probabilities is the tested energy range over the width of the measured signal. Therefore, in order to specify a kind of p value that is a reasonable indicator of the error probability, one must correct the local p value by the stated factor. Accounting for LEE by multiplying the local p value by that factor leads to the specification of a global p value that is taken to provide the most adequate characterization of the actual significance of measured data.

While global p values are calculated and stated in HEP-publications, the "official" discovery criteria are nevertheless bound to local p values, which constitute the experimental result that can be stated most straightforwardly. The very high discovery limit of 5σ, however, is motivated in part by the fact that LEEs of 100 or more are common in HEP and have to be absorbed in the discovery limits for local p values in order to have a sufficiently small risk of false discovery announcements. Looking specifically at the Higgs search, the LHC experiments were indeed sensitive to scalar particles within a wide mass range. Accounting for that range generated a LEE of roughly the order 100.

The described situation provides the background for the conflicting assessments of the December 2011 data. The adherents to the *theoretician's position* made the following points. First, a rigid 5σ criterion is not the most accurate way to account for the actual significance of our data. What one should do instead is evaluate the specific experimental situation. Specifically, one should quantify the actual LEE in the given case. Second, when doing so, it is not sufficient just to specify the tested parameter range. What we rather have to do is take into account all we know or believe to know about HEP already. In the given case, there were good theoretical reasons for expecting a very specific kind of scalar particle, namely a Higgs particle. Further, there

was good evidence already before December 2011 that the mass of the Higgs particle, if it existed at all, had to lie within a rather narrow mass range.[9] The signal found in 2011 indeed lay within that narrow mass range. The theoretical knowledge that the Higgs particle was the most likely candidate for a new scalar field in conjunction with experiments which constrained Higgs particles to specific mass values, according to the *theoretician's* perspective was reason enough to consider a reduced LEE that accounted only for the allowed Higgs mass region. Therefore, the global p value of the measured signal had to be stated to be much lower than naively assumed, which meant that a signal with a local p value of 4 σ could be taken (at least inofficially) as practically a discovery.

The *experimentalist's position* rejected this line of reasoning. Taking into account our trust in the existence of the Higgs particle while searching for that very particle seemed to violate the impartiality of the experimental process. Therefore, from the *experimentalist's* perspective, any reduction of LEE on that basis amounted to a distortion of the experimental result and had to be rejected. Thus the measured 4 σ result had to be considered no better than any of those 4 σ effects that had turned out to be due to statistical fluctuations in the past.

It is rather straightforward to understand the described difference of opinion in terms of a controversy between a rigidly frequentist and a partly Bayesian perspective. The rigid frequentist, that is the endorser of the *empiricist's position*, holds that any reference to prior probabilities is a distortion of the objective data and has to be rejected. The Bayesian, to the contrary, is ready to account for theory-based trust in a given theory by introducing prior probabilities for the truth or viability of that theory.

It is important to understand, however, that the endorser of the *theoretician's position* does by no means choose a fully Bayesian perspective on data analysis. As already pointed out in Sect. 1, HEP-physicists tend to agree on the goal to keep quantitative scientific data analysis inter-subjective: data analysis should be carried out in a way that makes the published quantitative results reproducible by anyone who knows experiment, data and the theory that is tested. A fully Bayesian perspective would amount to calculating the posterior probability for the Higgs hypothesis based on subjective priors by applying the Bayes formula $P(H|E) = P(E|H)\dfrac{P(H)}{P(E)}$. As discussed above, these priors would be informed by knowledge about the theoretical context. Nevertheless, they would be specific and non-generic quantitative expressions of subjective expectations by individual scientists regarding the theory's viability. A fully Bayesian approach thus would bluntly violate the condition of inter-subjectivity, which renders it just as unacceptable from the *theoretician's* as from the *experimentalist's* perspective. What the endorser of the *theoretician's* perspective subscribes to is accounting for prior probabilities in the interpretation of the data that leads up to specifying a discovery criterion. This part of data analysis is based on subjective considerations anyway, which means that a reliance on subjective priors looks much easier to accept. The prior probability for the Higgs hypothesis provides the basis for reducing LEE,

---

[9] This knowledge, as of December 2011, was partly based on calculations which showed that a Higgs with a higher mass would have had an observable effect already in earlier experiments based on virtual Higgs effects, even though no real Higgs particles would have been generated there.

which in turn can justify a less stringent criterion for calling a set of data a discovery of new physics.

This is roughly the line of reasoning presented in (Dawid 2015). In the next section, we want to go one step further and ask whether the *theoretician's position* can be understood and motivated from an epistemologically fully Bayesian perspective on HEP data analysis.

4 LEE from a Bayesian perspective

From a Bayesian perspective, one needs to express the frequentist results, that is local and global p values, in terms of $P(N|E)$: one asks for the probability of N given the collected data E. $P(N|E)/P(N)$ is the only legitimate basis for characterizing the Bayesian confirmation value of empirical data E. The problem is that neither the probabilities of signals stronger than the measured signal (which enter the calculation of the p-vale) nor any reference to LEE show up explicitly in $P(N|E)/P(N)$. Thus the question arises: can the Bayesian analysis explain why it makes sense to account for both in a frequentist context?

The focus of this section will be on the role of LEE. We will show that a satisfactory Bayesian understanding of reduced LEE can indeed be achieved once one is ready to assume that p values do have some relevance as an indicator of $P((N|E)$ at all.

The local p value expresses the probability that the measured number of Higgs candidate events or a higher number is produced at the given mass scale when the null hypothesis is true. In terms of conditional probabilities, one may express this as

$$p(E, N) = 2P(E_+|N) \tag{3}$$

where E_+ is the datum that the experiment has produced the number of events that was actually measured or a higher number. (The factor 2 enters because the Poisson distribution has an upper and a lower tail.) We can think of E_+ in terms of a counting device that counts events of a certain kind - that is, in our case, Higgs candidates in particle collisions - but cannot count beyond the number of events that have actually been counted: we thus don't know whether any events in excess of that number have occurred.

The connection that will be established in this section is between global p values and the probability $P(N|E_+)$. Of course, we must keep in mind that our actual data is E rather than E_+, which means that the most adequate verdict on N based on the data is given by $P(N|E)$. Specifying the relation between $P(N|E_+)$ to $P(N|E)$, however, requires further assumptions on the spectrum of alternatives to N. Some comments on the nature of that relation in the Higgs case will be given at the end of this section.

It is clear that the frequentist approach cannot account for those elements of Bayesian data assessment which are based on prior probabilities of the involved data and hypotheses. Thus, a Bayesian justification of the use of global p values in frequentist data analysis must amount to demonstrating that global p values provide the values closest to $P(N|E)$ (or, for the time being, $P(N|E_+)$) that can be reached after

discarding the direct contributions of the (subjective) prior probabilities of the theories involved.

We assume that a scalar particle is searched for within a given mass range and a significant signal E is found at at some mass scale. That is, the null hypothesis N that no scalar particles contribute to the measured signal has a very small p value. Apart from N, we distinguish between the Higgs hypothesis H and other new physics (to be denoted by $\neg H \wedge \neg N$) that allows for a scalar particle within the tested mass range. The three alternatives exhaust the space of possibilities. The Bayesian law and the law of total probability with respect to E_+ then give:

$$P(N|E_+) = P(E_+|N)\frac{P(N)}{P(E_+)} \tag{4}$$

$$P(E_+) = P(N)P(E_+|N) + P(H)P(E_+|H)$$
$$+ P(\neg N \wedge \neg H)P(E_+|\neg N \wedge \neg H) \tag{5}$$

The smallness of the p value implies that the data E_+ is highly improbable if N is true. Therefore, we can write

$$P(E_+) \cong P(T_{\neg N})P(E_+|T_{\neg N})$$
$$= P(H)P(E_+|H) + P(\neg N \wedge \neg H)P(E_+|\neg N \wedge \neg H) \tag{6}$$

Now we know the following:

1. The Higgs hypothesis H implies the existence of the Higgs particle.
2. Old data has excluded a Higgs particle with a mass outside a given parameter range R at $2\,\sigma$ confidence level.

Therefore the Higgs hypothesis in conjunction with old data implies (with some confidence) a scalar particle somewhere within the parameter range R of size R where we have $R/W = C_{LEE}$.

Here W is the signal width and C_{LEE} is the look elsewhere factor. Assuming (i) that the probability of the existence of a Higgs particle is uniformly distributed over the allowed mass region and (ii) that data E roughly corresponds to the expected significance of a Higgs signal (which means that $P(E_+|H) \cong P(E_-|H)$), we can thus write

$$P(E_+|H) \cong \frac{1}{2C_{LEE}} \tag{7}$$

so that

$$P(E_+) \cong P(H)\frac{1}{2C_{LEE}} + P(\neg N \wedge \neg H)P(E_+|\neg N \wedge \neg H) \tag{8}$$

This gives

$$P(N|E_+) \cong P(E_+|N)\frac{P(N)}{P(H)\frac{1}{2C_{LEE}} + P(\neg N \wedge \neg H)P(E_+|\neg N \wedge \neg H)} \tag{9}$$

At this point, we have to account for the scentists' assessment of the prior probability of H. Scientists expected that the Higgs hypothesis was likely true, which justifies attributing a high prior probability to it. In fact, we wouldn't even need a high absolute value for $P(H)$ for our purpose. It would be sufficient to assume that the $P(H)$ is considerably higher than $P(\neg N \wedge \neg H)$, the prior probability of the scenario that a scalar particle that is observable at the LHC must be accounted for by some other new theory. We thus assume

$$P(\neg N \wedge \neg H) << P(H) \tag{10}$$

Moreover, we have no reason to assume that other theories of new physics than the Higgs mechanism will lead to a scalar particle with a mass within the parameter range R. Therefore, we have

$$P(E_+ | \neg N \wedge \neg H) << \frac{1}{2C_{LEE}} \tag{11}$$

Assuming (10) and (11) or even making the weak assumption that $P(H)$ is not much smaller than $P(\neg N \wedge \neg H)$ in conjunction with a strong version of (11) suppresses the second term in the denominator of Eq. (9) against the first one, which leads to

$$P(N|E_+) \approx 2P(E_+|N)C_{LEE}\frac{P(N)}{P(H)} = p(E, N)C_{LEE}\frac{P(N)}{P(H)} \tag{12}$$

$\dfrac{P(N)}{P(H)}$ accounts for the subjective priors which influence $P(N|E_+)$. They are ignored in a frequentist approach. The remaining expression $p(E, N)C_{LEE}$ is the global p value.

We thus have shown that C_{LEE} is required for providing a characterisation of the significance of data E that comes as close to $P(N|E_+)$ as possible without explicitly introducing prior probabilities in the data analysis.

At this point we have to come back to the relation between $P(N|E_+)$ and $P(N|E)$. Without making assumptions on the most probable predictions beyond the null hypothesis N, nothing can be said about this relation. Based on the assumption that the Higgs hypothesis is the most probable alternative to N, however, and considering the case that it predicts roughly a signal of the significance of E, one can make an estimate. The easiest way to proceed is to replace the Poisson distribution by a continuous Gaussian probability distribution so that likelihoods correspond to probability densities. Roughly, $P(E_+|H) \cong 1/2$ is then of the same size as $P(E|H)$ (which is close to the maximum of the Gaussian curve). Therefore, when we replace E_+ by E in Eq. (4) and repeat the entire analysis up to Eq. (12), we find that the difference between $P(N|E_+)$ and $P(N|E)$ by and large corresponds to the difference between $P(E_+|N)$ and $P(E|N)$. If $P(E|N)$ is out on the tail of the Gaussian curve, $P(E|N)$ provides a significantly larger number than $P(E_+|N)$. The ratio between $P(E|N)$ and

$P(E_+|N)$ increases with higher significance of data E. [10] $P(N|E_+)$ thus tends to give a substantially smaller probability of the null hypothesis than $P(N|E)$. Understanding the p value in terms of a probability of the null hypothesis thus would significantly underestimate that probability. Still, disregarding the potentially strong impact of prior probabilities, the p value does offer a rough indication of the size of $P(N|E)$.

The quantitative difference between $P(N|E_+)$ and $P(N|E)$ offers one more conceptual reason for the very high general discovery criteria in HEP. On the basis of such a high general discovery criterion, it then makes sense to give a Bayesian argument for the deployment of a reduced LEE and therefore for a lower individual discovery criterion in special cases like the Higgs search. The reduced LEE factor that is extracted based on the *theoretician's* perspective thus does have a meaningful and consistent interpretation within a Bayesian framework.

There is no comparable way of making sense of the *experimentalist's* understanding of LEE within a Bayesian framework. From the *experimentalist's* perspective, the look elsewhere effect (to be called $\overline{LEE}$ as opposed to the theoretically informed LEE) is specified simply by considering the tested energy range $\bar{R}$ of size $\bar{R}$ without making any specific assumptions with respect to the prior probability of the Higgs hypothesis or any other theory. In order to make sense of $\overline{LEE}$ along the lines of the previous Bayesian analysis, we would need the relation

$$P(E_+|\neg N) \cong \frac{1}{2\bar{C}_{LEE}} \tag{13}$$

to hold, in analogy to Eq. (7). This relation could only be established if we assumed that, under the condition $\neg N$, the expectation value for the number of scalar particles to be found with a mass within the parameter range $\bar{R}$ was 1 and the probabilities of mass values were uniformly distributed within $\bar{R}$. In that case, accounting for $\overline{LEE}$ would indeed amount to specifying a global p value $p((E, N)$ in a way that differs from $P(N|E_+)$ only by disregarding the subjective priors for the spectrum of possible theories, just like in the *theoretician's* case. But there is no legitimation for that set of assumptions. From the theoretician's perspective, the assumption is untenable since it is at variance with the use of the reduced LEE in Eq. (7). And from the *experimentalist's* perspective, one knows nothing about the empirical implications of $\neg N$. There is no basis for assuming that the Higgs hypothesis H is more likely than any of the many potential unknown theories about new physics. Therefore the *experimentalist's* perspective offers no access to $P(E_+|\neg N)$ whatsoever. We conclude that it is impossible to make sense of the *experimentalist's* LEE within a Bayesian framework. In order to motivate the specification of LEE, we must endorse a t*heoretician's position* that allows us to specify prior probabilities for the existence of a scalar particle within the tested energy range.

We thus have seen that the *theoretician's position* not only implicitly employs Bayesian elements of reasoning but actually constitutes the only coherent understanding of LEE from an epistemically Bayesian perspective.

---

[10] At 5 σ, the p value is about 6 times the probability density. Accordingly, $P(N|E_+)$ would underestimate the probability of the null hypothesis by a factor 6.

5 LEE and non-empirical theory confirmation

Section 2 addressed a context where Bayesian elements of reasoning were considered in physical data analysis itself. Sections 3 and 4 discussed a case where one step in physical data analysis, the specification of LEE, could be modified based on considerations which involved Bayesian reasoning. In the present section, we will consider an even more central role of Bayesian reasoning. As it stands, those considerations may be called purely philosophical. Still, they do raise some questions that may become increasingly relevant for an overall physical understanding of theory confirmation.

Once again, we start with the observation that high energy physicists had a high degree of trust in the existence of a Higgs sector already before a Higgs candidate was empirically discovered. The trust in the Higgs mechanism was based on a combination of two arguments. First, it relied on a 'no alternatives argument' (NAA): there seemed no plausible way to give a coherent overall description of microphysics that did not involve a Higgs sector. NAA was applied to a very specific empirical question. One must clearly distinguish between the general hypothesis that some Higgs sector existed and more specific conjectures regarding the detailed form of that Higgs sector (whether scalars were fundamental or composite, whether there existed one or several scalar particles, whether or not there was low energy supersymmetry, etc.). NAA was used with respect to the first question. It was not used with respect to the second, where a wide variety of alternatives were—and still are—considered. NAA thus was—and always is—a strategy for assessing the prospects for the next step of theory building. It is not a strategy for announcing final or ultimate solutions.

In order to be convincing, NAA had to be supported by a second type of argument, to be called the meta-inductive argument from predictive success in the research program (MIA). Let us introduce MIA right away for the specific case of the Higgs hypothesis. The Higgs hypothesis is part of the standard model of particle physics. The motivation for developing the Higgs mechanism was to make gauge field theories like the standard model consistent with the observation of fermionic mass spectra and massive vector bosons. Now the standard model made a wide range of predictions apart from the predictions of a Higgs boson. All those other predictions were empirically confirmed between 1974 and 1994. Trust in the predictions of the standard model was, in one way or the other, always based on a no alternatives argument. Roughly, the chain of reasoning was the following. We have no plausible alternative to quantum field theory as a theory of interacting relativistic quantum objects at the electroweak scale. A fully consistent formulation of quantum field theory requires a renormalizable theory. Gauge field theory seems to be the only approach that brings about renormalizability. On the basis of the observed data, the standard model makes a certain prediction. This prediction, then, seems unavoidable based on the above line of reasoning.

Each time experimentalists succeeded in testing a standard model prediction, they confirmed that prediction, starting with neutral currents over symmetries of gauge couplings to new fermionic particles and specific vector Bosons for strong and (electro)weak interactions. This series of empirical successes amounted to strong confirmation of the standard model. At a meta-level, however, it also confirmed the hypothesis that theoretical claims in the standard model context which were made based on no-alternatives arguments had a high chance of being predictively success-

ful. Based on this meta-inductive argument (MIA), it seemed justified to take NAA very seriously in the case of the Higgs mechanism as well.

The question now arises as to what status can be attributed to NAA and MIA. In a Bayesian context, this question can be stated in a clear-cut way. Both NAA and MIA rely on observations at a meta-level. In the case of NAA, this is the observation, call it F_A^H, that no alternatives to theory H have been found. In the case of MIA it is the observation F_M^H that theories which belong to the same research program as H and to which NAA is applicable tended to be empirically successful in the past. The question is whether or not F_A^H and F_M^H increase the probability of the viability of theory H. If they do, they must be acknowledged as theory confirmation in a Bayesian sense. If they don't, neither NAA nor MIA are genuine arguments for the viability of H and both should be discarded.

Observations F_A^H and F_M^H don't lie within the intended domain of hypothesis H, i.e. they cannot be predicted by H. Therefore, they do not constitute evidence that is canonically understood to confirm H. In Dawid et al. (2015), it was shown that NAA nevertheless provides the basis for theory confirmation in a Bayesian sense.

In the following, I will briefly rehearse the basic line of reasoning that leads to establishing that NAA amounts to confirmation.

We assume that there exists a specific but unknown number k of possible scientific theories that satisfy reasonable scientificality conditions $\mathcal{C}$, explain the collected data $\mathcal{D}$ and predict the outcomes of a set of future experiments $\mathcal{E}$.

We then introduce the binary propositional variables T and F_A.

T takes the values

> T The hypothesis H is viable (that is, empirically adequate within a given regime).
> ¬T The hypothesis H is not viable.

Note that we talk about a theory's viability rather than truth. This emphasises the limited context of applicability of theory assessment. we are not interested in absolute truth but in a theory's predictive success within a certain regime. F_A takes the values

> F_A The scientific community has not yet found an alternative to H that fulfills $\mathcal{C}$, explains $\mathcal{D}$ and predicts the outcomes of $\mathcal{E}$.
> ¬F_A The scientific community has found an alternative to H that fulfills $\mathcal{C}$, explains $\mathcal{D}$ and predicts the outcomes of $\mathcal{E}$.

Since F_A does not lie in the intended domain of H, we introduce an additional variable Y that mediates the connection between T and F_A. Y has values in the natural numbers, and Y_k corresponds to the proposition that there are exactly k hypotheses that fulfil $\mathcal{C}$, explain $\mathcal{D}$ and predict the outcomes of $\mathcal{E}$.

Next, we make the following rather weak and plausible assumptions.[11]

A1. The variable T is conditionally independent of F_A given Y:

$$T \perp\!\!\!\perp F_A | Y \qquad (14)$$

---

[11] In Dawid et al. (2015), the set of assumption is a little more extensive since our assessment of the scientists' capacity of developing new theories is included in the analysis. For our present purposes, however, it is sufficient to state the following reduced set of assumptions.

A2. The conditional probabilities

$$f_{kj} := P(F_A|Y_k) \tag{15}$$

are non-increasing in k for all $j \in \mathbb{N}$ and non-decreasing in j for all $k \in \mathbb{N}$.
A3. The conditional probabilities

$$t_k := P(T|Y_k) \tag{16}$$

are non-increasing in k.
A4. There is at least one pair (i, k) with $i < k$ for which (i) $y_i y_k > 0$ where $y_k := P(Y_k)$, (ii) $f_{ij} > f_{kj}$ for some $j \in \mathbb{N}$, and (iii) $t_i > t_k$.

Assumption 1 implies that we would learn nothing new about the viability of H from our failure to find alternatives to H if we knew how many alternatives actually existed. Assumptions 2–4 amount to the requirement that there is at least a very mild dependence between the actual number of alternatives and both the probability of H being viable and the probability that no alternatives to H were found.

On that basis, the following theorem is proved in Dawid et al. (2015):

Theorem 1 *If Y takes values in the natural numbers $\mathbb{N}$ and assumptions* **A1** *to* **A4** *hold, then* F_A *confirms* T*, that is,* $P(T|F_A) > P(T)$.

F_A thus constitutes confirmation of H. Applying this statement to the case of the Higgs particle, we are therefore allowed to say: the observation that no alternatives to the standard model approach that included a Higgs mechanism were found increased the probability of the viability of the Higgs hypothesis. It is legitimate to talk about non-empirical theory confirmation of the Higgs particle already before its empirical discovery.

A similar argument can be made with respect to MIA (work in progress). In fact, there is an interesting mutual connection between MIA and NAA. We saw already that NAA enters MIA: in MIA, only theories to which NAA is applicable are considered. On the other hand, NAA without support from MIA remains very weak. While it can be formally established that F_A^H constitutes confirmation of H, the strength of this confirmation cannot be specified based on NAA alone. There are two possible explanations of F_A^H: either there actually are very few alternatives or scientists are just not clever enough to find them. F_A^H on its own, even after continued and intense unsuccessful searches for alternatives, can never distinguish between the two possible explanations and therefore can never strengthen one compared to the other. MIA, to the contrary, can distinguish between the two. Predictive success obviously cannot be explained by insufficient capabilities of the involved scientists. It can be explained, however, by a lack of possible alternatives: if very few possible alternatives to the theory that was developed exist, chances are good that the developed theory will be empirically successful. MIA therefore, by favouring the hypothesis of few possible alternatives over the hypothesis of insufficient capabilities of involved scientists, can establish the significance of NAA in a given research context. On that basis, a more thorough analysis of the theoretical context of H can strengthen F_A^H further and eventually lead

to the degree of confidence in the Higgs hypothesis that was prevalent among high energy physicists in the one or two decades before its empirical confirmation.

The Bayesian approach thus offers a justification for understanding NAA and MIA in terms of theory confirmation. This is of interest for the discussion of the previous Section in two ways. First, once one has chosen a Bayesian perspective with respect to NAA and MIA, any attempt of relating non-empirical to empirical theory confirmation seems to require a Bayesian perspective on the entire process. Therefore an analysis along the lines discussed in Sect. 4 seems essential for acquiring a coherent overall understanding of theory confirmation. Second, the use of priors for the Higgs hypothesis in the context of specifying LEE gains authority by our current analysis. Since the prior probabilities deployed in the context of empirical confirmation of the Higgs hypothesis themselves constitute results of a process of significant non-empirical theory confirmation based on NAA and MIA, they must not be treated as arbitrary speculations. Rather, they should be seen as the expression of early confirming steps which constitute an integral part of the overall process of theory confirmation. Scientists are well justified to take them seriously.

Of course, the problem remains that non-empirical theory assessment does not provide objective numbers based on the collected data. Thus, there remains the clear-cut distinction between likelihoods and p values which can be univocally extracted in a frequentist framework and the Bayesian probabilities which can not.

6 Conclusion

The three described contexts show three different roles of a Bayesian perspective in HEP data analysis. In the first case, we saw a technical utilization of Bayesian methodology within a solidly frequentist context of analysis.

The second case showed how a specific element of data analysis, the treatment of LEE, can find a clearer interpretation within a Bayesian epistemic framework. That interpretation proved capable of deciding for or against specific approaches towards LEE based on their compatibility with a Bayesian epistemic embedding. Nevertheless, the core data analysis remained based on the extraction of local p values, which is an entirely frequentist form of analysis.

The third case interpreted the assessment of the status of the Higgs hypothesis before the *discovery* of the Higgs particle as a form of theory confirmation by non-empirical evidence - that is by observations which do not lie within the intended domain of the confirmed hypothesis. This concept makes sense only within a Bayesian framework and therefore suggests a fully Bayesian perspective. Any coherent characterization of the transition from a phase of non-empirical theory confirmation to the discovery of corresponding particles must be based on a Bayesian overall perspective as well, which speaks in favour of an epistemically Bayesian embedding of frequentist data analysis.

The second and third case arise in connection with the Higgs discovery because of the high degree of theory based trust in the Higgs hypothesis. They are of wider interest however, in the light of recent developments which indicate a rapprochement of HEP and cosmology. As exemplified by the recent scientific debate on empirical

confirmation for cosmological inflation, situations where empirical data is available but inconclusive may become a standard situation when assessing the viability of modern theories in HEP and cosmology. A coherent understanding of this kind of situation arguably may be based on understanding the relation between empirical and non-empirical theory confirmation which, in turn, seems to suggest an epistemically Bayesian framework of analysis. The Higgs hypothesis seems a good test case to that end.

References

ATLAS Collaboration, Aad, G., Abandian, T., et al. (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. arXiv:1207.7214.

CMS Collaboration, Chatrchyan, S., Khchatryan, V. et al. (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. arXiv:1207.7235.

Cousins, R. D. (1995). Why isn't every physicist a Bayesian? *American Journal of Physics*, *63*, 398.

Dawid, R., Hartmann, S., & Sprenger, J. (2015). The No Alternatives Argument. *The British Journal of Philosophy of Science*, *10*, 1093. doi:10.1093/bjps/axt045.

Dawid, R. (2015). Higgs discovery and the look elsewhere effect. *Philosophy of Science*, *82*(2), 76–96.

Howson, C. (2000). *Hume's Problem: Induction and the Justification of Belief*. Oxford: Oxford University Press.

Lyons, L. (2013). Bayes and frequentism. arXiv:1301.1273

Mayo, D. (1996). *Error and the growth of experimental knowledge*. Chicago: University of Chicago Press.

Mayo, D., & Spanos, A. (2006). Severe testing as a basic concept in a Neyman–Pearson philosophy of induction. *British Journal for the Philosophy of science*, *57*, 323–357.

Norton, J. D. (2008). Ignorance and Indifference. *Philosophy of Science*, *75*(2008), 45–68.

Read, A. L. (2002). Presentation of search results: The CLS technique. *Journal of Physics*, *G28*, 2693–2704.

Robertson, R. G. H., & Knapp, D. A. (1988). Direct measurements of neutrino mass. *Annual Review of Nuclear and Particle Science*, *38*, 185–215.

Staley, K. (2004). *The Evidence for the Top Quark: Objectivity and bias in collaborative experimentation*. Cambridge: Cambridge University Press.