Modelling Non-Empirical Confirmation

Richard Dawid*

October 17, 2015

Abstract

The paper provides a presentation and motivation of the concept of non-empirical theory confirmation. Non-empirical theory confirmation is argued to play an important role in the scientific process that has not been adequately acknowledged so far. Its formalization within a Bayesian framework demonstrates that non-empirical confirmation does have the essential structural characteristics of theory confirmation.

1 Introduction

The canonical view of the scientific process understands theory confirmation in terms of a direct confrontation of a theory's predictions with empirical data. A scientific theory is expected to make testable empirical predictions. If the relevant collected data agrees with those predictions, the data confirms the theory. If the data disagrees with the predictions, the theory gets disconfirmed.

One may view this understanding in terms of a technical definition of theory confirmation, which would render it immune against criticism. It may be argued, however, that the concept of confirmation should account for the scientists' actual reasons for taking a theory to be well-established as a viable description of a given aspect of the observed world. Endorsing that aim, one may question a given understanding of theory confirmation by comparing it with the scientists' actual attitude towards their theories.

The latter view is the point of departure chosen in the present article. It is assumed that the concepts deployed by the philosophy of science for

---

*LMU Munich – Munich Center for Mathematical Philosophy – http://homepage.univie.ac.at/richard.dawid/ – richard.dawid@univie.ac.at.

modelling scientific reasoning should offer a characterization of the actual structure of scientific reasoning - and should be measured by that standard. On that account, however, a closer look at actual science throws the adequacy of the canonical understanding of theory confirmation into doubt. In many scientific fields, confirmation in the canonical sense described above is not the only basis for an assessment of a theory's status. Three interrelated issues arise, which render an exclusive focus on empirical confirmation insufficient. They shall be briefly sketched in the following.

1: In historical fields of research, scientists often face a conjunction of two problems. First, the general character of scientific hypotheses in those fields often makes it difficult to extract specific and quantitative predictions from them. Second, and maybe even more troubling, those scientific fields fields often deal with empirical situations where most of the empirical record has been irretrievably lost to natural decay or destruction during the periods that lie between that events under investigation and the time of inquiry. Moreover, even of the data that would be available in principle, it is often only possible to collect a haphazard and arbitrary subset.[1] Anthropologists searching for early human traces, to give one example, cannot search specifically for the missing link they are most interested in but must be content with whatever new material their excavations provide. The two described conditions in conjunction create a situation where empirical confirmation remains patchy and, on its own, does not provide a stable foundation for assessing the probability that a theory is trustworthy. External characteristics of the theory and the research field therefore play an important role in that assessment.

More specifically, if various conflicting hypotheses aim at explaining the same available data, abductive forms of reasoning are deployed, which depend on understanding whether or not one of the theories seems substantially more plausible than the others. One important issue that must be addressed in such cases is the question whether and if so on what grounds it makes sense to assume that those theories that have been developed cover the spectrum of possible plausible theories on the issue. Only if that is the case does it make sense to trust the most plausible of the known theories. Trust in a theory thus is instilled based on a combination of assessments of the spectrum of known alternatives and some induced understanding of the presumptive spectrum of unconceived alternatives.

2: A similar issue arises in the case of micro-physical theories that conjecture the existence of unobservable physical objects like atoms, quarks or, to mention an important recent example, the Higgs particle. In those cases, an-

---

[1] For an instructive philosophical perspective on historical sciences, see (Turner 2007).

nouncing a discovery amounts to endorsing all empirical implications of the discovered object, whether or not they have been empirically tested yet. A discovery therefore has profound consequences in high energy physics. Once a particle has been discovered in one experiment, background calculations in all future experiments factor in all empirical implications the particle has within the established theoretical framework. The question is, however, on what basis scientists can be so confident that no unconceived alternative could account for the data collected without having the same further empirical implications as the known theory. The answer is that scientists cannot make that assessment without relying on observations about the overall research process. They need to make an assessment as to whether or not an alternative seems likely to show up based on their understanding of the overall conceptual context and whether assessments of that kind have turned out reliable in the past. In other words, the declaration of a discovery of a new object in microphysics relies on considerations very similar to those which lead scientists towards endorsing a theory in palaeontology or other non-formalized historical sciences.

3: Finally, since the 1980s high energy physicists and cosmologists have shown an increasing readiness to invest a high degree of trust in empirically unconfirmed theories. Theories like string theory or cosmic inflation are taken by many as important steps towards a deeper understanding of nature even though those theories so far have no (in the case of string theory) or only inconclusive (in the case of inflation) empirical confirmation. Once again it turns out that the reasons responsible for that trust are of a very similar kind as those at play in the previously discussed contexts.

Unlike in the previously discussed cases, the extent to which exponents of empirically unconfirmed theories in fundamental physics consider their theory well-established has led to a highly controversial debate on the scientific legitimacy of the involved strategies of theory assessment. In this light, the case of scientific trust in empirically unconfirmed theories turns the question of an adequate understanding of the concept of confirmation from a mainly philosophical issue into a question of high significance for the further evolution of fundamental physics.

All three discussed scientific contexts suggest that a perspective that focusses entirely on the agreement between a theory's predictions and empirical data is insufficient for acquiring an adequate understanding of the reasons why scientists trust a scientific theory.

In the following, I will present a widened concept of theory confirmation that, as I will argue, comes closer to that goal.

Two basic guidelines will determine the layout of the presented approach. On the one hand, as already pointed out above, the discussion will be guided by the idea that the concept of confirmation should provide a basis for understanding the degree of trust scientists have in a theory. On the other hand, however, the empirical character of science, that is the connection between confirmation and observation, shall not be abandoned. So, while the approach to be presented is non-canonical, it will be argued to remain true to core principles of scientificality.

2 The Setup

2.1 What is Non-empirical Confirmation?

The canonical view as it is presented in accounts from classical hypothetico-deductivism to most readings of Bayesian confirmation (see e.g. Bovens and Hartmann 2003, Howson and Urbach 2006) constrains confirmation to observations within the theory's intended domain. Only the agreement between a theory's predictions and empirical data constitutes confirmation of that theory. We will call this form of confirmation "empirical confirmation" because it is based on empirical testing of the theory's predictions. Our question will be: which kinds of consideration beyond the limits of empirical confirmation may in principle be understood as contributions to theory confirmation? More specifically, we'll search for a form of "non-empirical" theory confirmation that can account for those considerations that have been argued above to be crucial in a number of contexts for instilling trust in a scientific theory.[2] At the same time, however, we want to retain the crucial role of observation and stay as close as possible to the mechanism of empirical confirmation. Confirmation should remain true to the basic principles of scientific reasoning.

In order to guarantee a grounding of confirmation in observation, we introduce elements of empiricist reasoning at two distinct levels. First, we understand trust in a theory in terms of the theory's empirical predictions rather than in terms of truth. If a scientist trusts a theory, she believes that the theory's predictions in its characteristic regime, if tested, will get empirically confirmed. If a theory's predictions in its characteristic regime are indeed in agreement with all possible data, the theory shall be called empirically viable. Non-empirical confirmation thus amounts to an increase of trust in the theory's empirical viability. Note that this understanding of confirmation in a certain respect stays closer to an empiricist understanding

---

[2] The concept was first laid out in (Dawid 2006) and then further developed in (Dawid 2013). A Bayesian formalization of one argument of non-empirical confirmation was given in (Dawid, Hartmann and Sprenger 2015).

than concepts of empirical confirmation that are based on truth probability. By avoiding reference to truth, we block the possibility that theories which have no empirical implications get non-empirically confirmed. Trust in theories that have no empirical implications is trivial on our account and cannot be increased by any means. Therefore, confirmation of non-predictive theories cannot occur.

Second, we will require that confirmation be based on some observations about the world beyond the theory and its endorser. The mere fact that a consideration contributes to a person's subjective belief in a theory's viability does not justify calling that consideration non-empirical confirmation on our account. For example, the fact that some scientists trust elegant theories does not imply that a theory's elegance constitutes non-empirical confirmation.

Being based on observations about the world is a fairly vague requirement, however. Which kind of relation between observation and confirmation do we require? One might follow various strategies in this regard. One plausible guideline, which we will follow, is a structural similarity between empirical and non-empirical confirmation.

We will introduce the following fairly specific definition of non-empirical confirmation. Non-empirical confirmation is based on observations about the research context of the theory to be confirmed. Those observations lie within the intended domain of a meta-level hypothesis about the research process and, in an informal way, can be understood to provide empirical confirmation of that meta-level hypothesis. The meta-level hypothesis, in turn, is positively correlated with the probability of the truth or viability of the scientific theory under scrutiny.

This may seem like a fairly complicated and arbitrary construction at first glance. However, it has a number of considerable merits. Most significantly, non-empirical confirmation of the suggested kind turns out to work as a reconstruction of the most conspicuous lines of reasoning that do generate trust in a scientific theory beyond the limits of empirical confirmation.

Second, non-empirical confirmation in the suggested sense can be understood in terms of an extension of the basis of observational evidence for a theory. The mechanisms of connecting observations to the overall conceptual framework remain the same as in the case of empirical confirmation. Confirmation is still based on comparing predictions with observations, but that comparison may play out at the meta-level of analysing the research process within which the theory is embedded rather than at the ground level of the theory's subject matter.

Third, and directly related to point 2, non-empirical confirmation of the described kind resembles empirical confirmation in being symmetric between confirmation and dis-confirmation. Observations at the meta-level may equally support and speak against a theory's viability, depending on whether or not they agree with the predictions of the meta-level hypothesis. The correlation between observational input and confirmation/dis-confirmation thus works along very similar lines as in empirical confirmation.

2.2 Towards a Formalized Model

Conformation today is mostly understood in Bayesian terms. In this light, we will analyse the nature of non-empirical confirmation from a Bayesian perspective. It will turn out that a probabilistic approach is particularly suitable for characterizing the way non-empirical confirmation works.

In Bayesian terms, an increase of trust in a theory's viability is expressed as an increase of the subjective probability that the theory is viable. As already discussed, our use of probabilities of viability constitutes a deviation from canonical Bayesian epistemology , which is based on truth probabilities.

We introduce the proposition T that a theory H is empirically viable (consistent with all empirical data) within a given context. Let us first consider the case of empirical confirmation. We take H to be confirmed by empirical data E iff $P(\mathrm{T}|\mathrm{E}) > P(\mathrm{T})$, that is if the subjective probability of the viability of H is increased by E. If E lies within the extended domain of H, one can deduce a probability of E from H and a set of initial conditions specified based on other observations. A high probability of E then justifies

$$P(E|T) > P(E) \tag{1}$$

which implies that E confirms H due to Bayes' formula

$$\frac{P(T|E)}{P(T)} = \frac{P(E|T)}{P(E)} \tag{2}$$

Now our goal is to replace E by some observations F^X that are not in the intended domain of H. In other words, H in conjunction with knowledge about initial conditions for the system described by H does not provide any information on the probability of F^X. Nevertheless F^X should imply $P(\mathrm{T}|\mathrm{F}^{\mathrm{X}}) > P(\mathrm{T})$.

Further, we want this probability increase to be induced via a new hypothesis Y that lives at the meta-level of theory assessment and is positively correlated with both F and T. Moreover, F^X should be in the intended domain of Y, that is, implications for F^X can be extracted from hypothesis Y.

In principle, one might try to find a specific variable Y^X for each type of non-empirical observation F^X. However, we shall pursue a different strategy and specify one Y that will be tested by various forms of F^X. This strategy has two advantages. First, it turns out to work well with respect to the three most conspicuous candidates for non-empirical theory confirmation to be found in science. And second, it allows for a more coherent overall understanding of the way the arguments of non-empirical confirmation mutually support each other.

So what would be a plausible candidate for Y? It is helpful to think about this question by looking at the most straightforward candidate for an argument of non-empirical theory confirmation: the no alternatives argument (see Dawid, Hartmann and Sprenger 2015). Let us, for the time being, continue the analysis within the framework of this specific argument. Later, we will return to a more general perspective.

3 The No Alternatives Argument

Let us assume that we make the following observation F^A: scientists have looked intensely and for a considerable time for alternatives to a known theory H that can solve a given scientific problem but haven't found any. This observation may be taken by us as an indication that the theory they have is probably viable.

Clearly, this kind of reasoning plays an important role in generating trust in some empirically unconfirmed or insufficiently confirmed theories in science. As mentioned above, a specific reconstruction of a phenomenon or object in anthropology or other historic sciences gains credibility if the case can be made that no other plausible reconstruction has been found. Most high energy physicists believed in the existence of a Higgs particle even before its discovery in 2012 because no satisfactory explanation of the mass spectrum of elementary particles that did not rely on some kind of Higgs particle had been found.

We call this kind of reasoning the no alternatives argument (NAA) (Dawid, Hartmann and Sprenger 2015). In the following, we give a Bayesian reconstruction of NAA. In the case of NAA, we can easily identify the most natural candidate for Y: to the extent F^A increases the probability of the viability of H, it arguably does so by supporting the hypothesis that there in fact are no or very few possible scientific alternatives to H. NAA thus involves an inference from an observation F_A on the alternatives discovered to a statement Y on the actual number of possible alternatives. Y thus is a statement on the limitations on the number of possible scientific theories on a subject. In order to to make sense of this, we need to specify the framework more clearly. Let us assume a theory H that is built to account for empirical data $\mathcal{D}$. We now assume that there exists a specific but unknown number i of possible scientific theories (i.e. theories that satisfy a set of scientificality constraints $\mathcal{C}$) which are compatible with the existing data $\mathcal{D}$ and give distinguishable predictions for the outcome of some relevant set $\mathcal{E}$ of future experiments.

Here, $\mathcal{D}$ specifies the empirical status quo. Possible scientific theories on the subject must be consistent with the relevant available data $\mathcal{D}$. In the most straightforward cases, H can be shown either to predict data $\mathcal{D}$ or at any rate to be consistent with it. There are also more difficult cases (like e.g. string theory) where consistency of H with data $\mathcal{D}$ has not been established but is considered plausible. Obviously, a situation of the latter type generates a comparably lower prior probability $P(T)$. Still non-empirical confirmation can work on that basis as well.

$\mathcal{C}$ specifies what counts as a scientific theory. Only those theories that meet the scientificality-conditions $\mathcal{C}$ count as possible theories. Scientificality-conditions are themselves volatile to a given degree and may change in time. Note, however, that our formal argument does not rely on a precise specification of the scientificality-conditions. All we need is the assumption that scientists apply a set of scientificality-conditions that contains a viable core that can be satisfied by theories that are empirically viable with respect to the future experiments $\mathcal{E}$.

Having introduced a framework for scientific theory building, we still need to specify a way of individuating theories. We need to decide up to which point we still speak of one theory and when we start talking about two different theories.

Generally speaking, we individuate theories by their predictive implications with respect to future experiments $\mathcal{E}$. Theories which give the same predictions (or the same range of predictions under variation of their parameter values) count as one theory. The reason for choosing this approach is that we are mainly interested in empirical predictions. Trust in a theory, from our perspective, is justified if the theory ends up making correct predictions. Since we only talk about empirical viability and not about truth, questions related to the spectrum of empirically equivalent theories lie beyond the scope of our analysis.

The specific form of criteria for theory individuation depends on the kind of predictions one is interested in. Therefore, we don't prescribe those criteria in detail. Scientists implicitly select them in dependence on the level at which they trust their theory's predictions. Let us explicate this by looking at the example of the empirical testing of the Higgs model. The Higgs model is a theoretical concept that can explain why elementary particles have masses. It predicts the existence of a scalar particle with a mass that lies within a certain range of possible values. Physicsts had a high degree of trust in the existence of the Higgs particle already long before the particle's discovery in 2012. Let us now assume that, before 2012, some physicist wanted to predict that the Higgs-particle existed and had precisely mass M_1. This 'theory', let us call it H_1, would have been distinct from any other exemplification of the Higgs model that predicted a different mass for the Higgs particle. In order to count the alternatives to H_1 one would have had to count each of these variations as an individual 'theory' H_n and thus would have got an infinite number of possible alternatives to H_1. Given that physics before 2012 did not offer arguments for the precise Higgs mass, it would have been clear that one could not trust H_1 or its predictions. Individuating theories based on specific Higgs masses thus would have been an inadequate basis for deploying NAA with respect to the Higgs hypothesis.

Since there was no basis for predicting the precise Higgs mass before 2012, physicists were most interested in the question as to whether the Higgs particle exists at all without specifying its precise mass. They were interested in the general viability of the Higgs hypothesis as a theoretical mechanism that could explain particle masses and implied the existence of at least one scalar particle – and they were quite confident about the viability of the Higgs mechanism even in the absence of empirical data. When analysing this situation, an assessment of possible alternatives to the Higgs hypothesis must not count different mass values as different theories. Even the specification of the Higgs model beyond its core structure (by introducing addditional scalars, a constituent structure of the Higgs particle, etc.) would not have counted as a different theory at this level. Only substantially different approaches to mass generation which did not rely on a scalar field would have counted as alternatives to the Higgs hypothesis. The fact that one had not found any convincing alternatives at this level gave reason for having trust in the viability of the Higgs hypothesis even in the absence of empirical confirmation. The level of theory individuation used in NAA had to correspond to this line of reasoning.

Having thus clarified the framework for specifying the number i of possible alternatives to theory H, we can now proceed to the proof that NAA amounts to confirmation of H in Bayesian terms based on a set of very plausible assumptions.

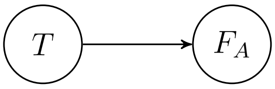

Figure 1: The Bayesian Network representation of the two-propositions scenario.

4 Formalizing the No Alternatives Argument

We introduce the binary propositional variables T and F_A, already encountered in the previous section.[3] T takes the values

- T The hypothesis H is empirically viable.
- $\neg\mathrm{T}$ The hypothesis H is not empirically viable.

and F_A takes the values

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

We would now like to explore under which conditions $\mathrm{F_A}$ confirms H, that is, when

$$P(\mathrm{T}|\mathrm{F_A}) > P(\mathrm{T})\,. \tag{3}$$

We then introduce variable Y that mediates the connection between T and F_A. In the previous section, we characterized Y in general terms as a statement about limitations to the number of possible alternatives to theory H. In our formalization, we are more specific. Y has values in the natural numbers, and $\mathrm{Y_k}$ corresponds to the proposition that there are exactly k hypotheses that fulfil $\mathcal{C}$, explain $\mathcal{D}$ and predict the outcomes of $\mathcal{E}$.

The value of F_A—that scientists find/do not find an alternative to H—does not only depend on the number of available alternatives, but also on the relation between the difficulty of the problem and the capabilities of the scientists. Call the variable that captures this factor S, and let it take values in the natural numbers, with $\mathrm{S_j} := \{S = \mathrm{j}\}$ and $d_j := P(\mathrm{S_j})$. The higher the values of S, the more difficult the problem is to solve for the scientists.[4] It is

---

[3] The presentation of this section is largely taken from (Dawid, Hartmann and Sprenger 2015)

[4] For the purpose of our argument, it is not necessary to assign a precise operational meaning to the various levels of S. It is sufficient that they satisfy a natural monotonicity assumption with regard to their impact on F_A—see condition **A3**.

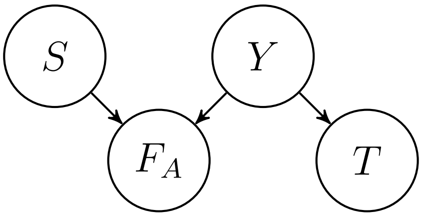

Figure 2: The Bayesian Network representation of the four-propositions scenario.

clear that S has no direct influence on Y and T (or vice versa), but that it matters for F_A and that this influence has to be represented in our Bayesian Network.

We now list five plausible assumptions that we need for showing the validity of the No Alternatives Argument.

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

$$T \perp\!\!\!\perp F_A | Y \tag{4}$$

Hence, learning that the scientific community has not yet found an alternative to H does not alter our belief in the empirical adequacy of H if we already know that there are exactly k viable alternatives to H.

A2. The variable D is (unconditionally) independent of Y:

$$D \perp\!\!\!\perp Y \tag{5}$$

Recall that D represents the aggregate of those context-sensitive factors that affect whether scientists find an alternative to H, but that are not related to the number of suitable alternatives. In other words, D and Y are orthogonal to each other by construction.

These are our most important assumptions, and we consider them to be eminently sensible. Figure 2 shows the corresponding Bayesian Network. To complete it, we have to specify the prior distribution over D and Y and the conditional distributions over F_A and T, given the values of their parents. This is done in the following three assumptions.

A3. The conditional probabilities

$$f_{kj} := P(\mathrm{F_A}|\mathrm{Y_k}, \mathrm{D_j}) \tag{6}$$

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

The (weak) monotonicity in the first argument reflects the intuition that for fixed difficulty of a problem, a higher number of alternatives does not decrease the likelihood of finding an alternative to H. The (weak) monotonicity in the second argument reflects the intuition that increasing difficulty of a problem does not increase the likelihood of finding an alternative to H, provided that the number of alternatives to H is fixed.

A4. The conditional probabilities

$$t_k := P(\mathrm{T}|\mathrm{Y_k}) \tag{7}$$

are non-increasing in k.

This assumption reflects the intuition that an increase in the number of alternative theories does not make it more likely that scientists have already identified an empirically adequate theory.

A5. There is at least one pair (i, k) with $i < k$ for which (i) $y_i \, y_k > 0$ where $y_k := P(\mathrm{Y_k})$, (ii) $f_{ij} > f_{kj}$ for some $j \in \mathbb{N}$, and (iii) $t_i > t_k$.

In particular, this assumption implies that $y_k < 1$ for all $k \in \mathbb{N}$ because otherwise, a pair satisfying (i) could not be found.

With these five assumptions, we can show that (For a proof, see Dawid, Hartmann and Sprenger 2015):

Theorem 1. *If Y takes values in the natural numbers $\mathbb{N}$ and assumptions* **A1** *to* **A5** *hold, then* $\mathrm{F_A}$ *confirms* T*, that is,* $P(\mathrm{T}|\mathrm{F_A}) > P(\mathrm{T})$.

$\mathrm{F_A}$ thus confirms the empirical viability of H under rather weak and plausible assumptions.

5 The Meta-Inductive Argument

So NAA formally constitutes confirmation. The question remains, however, how significant that confirmation is. The problem is that we have two possible explanations of F_A. F_A may be explained by the fact that there are no or very few possible alternatives to H. However, it might also be explained by a statement of type S: scientists are not clever enough to find the complicated alternatives that are possible. F_A cannot distinguish between those two kinds of explanation. If it is our prior assessment that an explanation of type S is far more likely to apply than an explanation of type Y, even the most powerful observation F_A could not alter this assessment. Therefore, if we start with very low priors for low number Y_ks and high priors for the hypothesis that scientists are not clever enough to find most of the alternatives, F_A won't strongly increase probabilities for low number Y_ks and therefore won't provide significant confirmation of H.

In order to turn NAA into a significant argument, we therefore need a second line of reasoning that allows us to distinguish between S and Y and, on that basis, can establish considerable probabilities of low number Y_ks which can then serve as a basis for significant confirmation of H by NAA.

This brings us to the second argument of non-empirical confirmation, the meta-inductive argument (MIA). The meta inductive argument is based on the observation F_M that those theories in the research field that satisfy a given set of conditions K (note that these are not the scientificality conditions C but may be considerably more restrictive) have shown a tendency of being viable in the past.

A meta-inductive step leads directly from F_M to inferring a high probability $P(T|F_M)$. However, in order to use MIA as support for NAA, it is helpful once again to use the statements Y as an intermediary. In order to do so, we have to assume a stronger connection between empirical viability and the number of alternatives. The simplest and most straightforward assumption would be that the theory found by the scientists is a random pick from the set of empirically distinguishable possible alternatives. This means that a theory's chances of being viable is $P(T) = 1/i$. Based on this model one can understand our subjective probability $P(T)$ in terms of our assessment of the probabilities of numbers of alternatives. For the simple model introduced above we get

$$P(T) = \sum_k P(Y_k)P(T|Y_k) = \sum_k \frac{1}{k} P(Y_k). \tag{8}$$

On that basis, if one observes a certain success rate of theories in a research field that satisfy conditions K, a frequentist analysis enforces substantial probabilities for Y_ks with low k. To give an example, let us assume that we observe a success rate of 50% of theories that satisfy K. A simple calculation shows that, based on our model and frequentist data analysis, we must attribute a probability of 1/3 or higher to the hypothesis ($k = 1 \vee k = 2$). MIA therefore generates assessments of $P(Y_k)$ which can then serve as priors in NAA.

MIA thus strengthens explanation Y of F_A and weakens explanation S correspondingly: if scientists have been so successful in finding viable theories in the past, it seems less plausible to assert that they are not clever enough for doing the same this time. Therefore, MIA can turn NAA into a method of significant confirmation.

One important worry may arise even if MIA and NAA look strong: it is not a priori clear whether the empirically unconfirmed theory that is evaluated is sufficiently similar in relevant respects to earlier successful theories to justify meta-inductive inference from the viability of those earlier theories to the viability of the present one.

Now it may happen that the theory under evaluation is so closely related and the problems addressed are so similar to earlier cases that there just seems no plausible basis for that worry. The Higgs hypothesis is an example of such a scenario. It is so deeply immersed in standard model physics that it would be quite implausible to argue that physicists understood the other problems raised with respect to the standard model sufficiently well but were not up to the complexity of the Higgs case.

In a similar vein, certain principles of reconstructing species from scarce excavated evidence may be applicable in many fairly similar individual cases. If such a strategy has proved successful in a number of cases, this may be sufficient for trusting the reliability of the method in similar cases in the future, provided the method offers the only known plausible reconstruction of the given species.

In cases like those referred to above, NAA and MIA in conjunction can be sufficient for generating a high degree of trust in a theory. The trust in the Higgs mechanism was indeed generated largely by those two arguments: the understanding that no convincing alternative to the Higgs mechanism was forthcoming and the observation that standard model physics had turned out predictively extremely successful whenever it had been tested.

6 Unexpected explanatory interconnections

There are other cases, however, where differences between previous theories and the new one with respect to the nature or the complexity of the core problems are very significant. String physics is a particularly good example of such a situation. Though string theory stands in the tradition of previous high energy physics, it clearly is a far more complex and difficult theory than any of its forebears. Thus it may easily be argued that, while scientists were clever enough to understand the spectrum of possibilities in the case of standard model physics, they are not up to the task with respect to string physics.

In cases like this, a third argument is required in order to turn NAA+MIA into a convincing line of reasoning. Arguably, the most effective argument of this kind is the argument of unexpected explanatory interconnections (UEA). The observation F_U on which this argument is based is the following. Theory H was developed in order to solve a specific problem. Once H was developed, physicists found out that H also provides explanations with respect to a range of problems which to solve was not the initial aim of developing the theory.

The argument is structurally comparable to the argument of novel empirical confirmation: a theory that was developed in order to account for a given set of empirical data correctly reproduces data that had not entered the theory's construction process. UEA is the non-empirical "cousin" of novel confirmation. Instead of successful novel empirical predictions, the theory provides unexpected explanatory interconnections that do not translate into successful empirical predictions.

The most prominent example of UEA in the context of string theory is the microphysical derivation of the black hole entropy area law in special cases of black holes. String theory was not developed for providing this derivation. More than two decades after string theory was proposed as a theory of all interactions, Vafa and Strominger (1996) succeeded in providing it.

UEA fits well into the category of non-empirical confirmation because it can be read as an argument for limitations to underdetermination just like NAA and MIA. The line of reasoning in the case of UEA is the following. Let us assume a set of n seemingly unrelated scientific problems in a research field. Let us further assume that there is a number i of possible alternative solutions to one of those problems. If the number of possible solutions to a specific problem is typically much higher than n, we have no reason to expect that a random solution to one problem will solve other problems as well. If we assume, however, that i is typically substantially smaller than n, we may plausibly assume that consistent strategies for solving one individual problem will typically be applicable to a number of problems. The reason for this is that we know that there is one theory, the true theory, that does solve all n problems. Therefore, in the extreme case that there is only one consistent solution to the problem we look at, all problems must be solved by that theory. Inversely, the observation that the theory that was developed for solving the given problem turns out to answer a number of other open questions as well can be taken as an indicator that there probably are very few possible different solutions to the given problem. From that consideration, once again, there follows a reasonably high probability that the given theory is viable.

UEA is of particular importance in contemporary fundamental physics, where theory building gets extremely difficult and UEA can provide a check as to whether or not physicists are capable of dealing with the overall set of problems they face in a way that goes beyond limited puzzle solving with respect to individual problems.

7 Conclusion

What is the status on non-empirical theory confirmation? As already emphasised in the introduction, non-empirical confirmation is an extension of empirical confirmation with a widened arsenal of conceptual tools but similar basic patterns of reasoning. It is secondary to empirical confirmation for two reasons. First, non-empirical confirmation is understood as a tool for establishing a theory's viability. Viability however, as defined in Section 2, is based on the theory's agreement with empirical data. Therefore, the perspective of eventual empirical testing is always in the background and, once conclusive empirical testing can be achieved, will in all cases make a stronger case for the theory's viability than non-empirical confirmation ever could. Second, the significance of non-empirical confirmation crucially relies on MIA. MIA, however, as described in Section 5, is based on empirical confirmation somewhere else in the research field. Non-empirical confirmation therefore can only work properly as long as empirical confirmation can be achieved somewhere in the research field.

Non-empirical confirmation is closely linked to a probabilistic view on confirmation. To a philosopher who denies that confirmation has anything to do with attributing a probability to a theory's viability or truth, non-empirical confirmation will look empty. On the other hand, to a philosopher who acknowledges a probabilistic basis of confirmation, it seems difficult to deny that non-empirical confirmation exists. From that perspective, the core question becomes how significant non-empirical confirmation can be. This paper offered some indications as to how a formal argument in favour of the significance non-empirical confirmation could be developed.

It has to be emphasised that a general argument for the legitimacy of non-empirical confirmation by no means implies that each individual deployment of non-empirical confirmation is convincing. There are cases in science, some of which have been mentioned in this paper, where the actual strength and influence of non-empirical arguments for a theory's viability is indeed striking. There are many others where understanding the strengths and weaknesses of the use of non-empirical confirmation requires careful analysis. I suggest that a probabilistic account of non-empirical confirmation provides a promising framework for carrying out that kind of analysis in a fruitful way.

References

Bovens, L., and Hartmann, S. [2003]: *Bayesian Epistemology*, Oxford: Oxford University Press.

Dawid, R. [2006]: ‘Underdetermination and Theory Succession from the Perspective of String Theory’, *Philosophy of Science*, **73**, pp. 298–322.

Dawid, R. [2009]: ‘On the Conflicting Assessments of the Current Status of String Theory’, *Philosophy of Science*, **76**, pp. 984–96.

Dawid, R. [2013]: *String Theory and the Scientific Method.* Cambridge: Cambridge University Press.

Dawid, R., Hartmann, S. and Sprenger, Jan [2015]: ’The No Alternatives Argument’, British Journal for the Philosophy of Science 66(1), 213-234.

Howson, C., and Urbach, P. [2006]: *Scientific Reasoning: The Bayesian Approach,* third edition, La Salle: Open Court.

Strominger, A. and Vafa, C. [2006]: ’Microscopic Origin of the Bekenstein Hawking Entropy’, Phyiscs letters B379, 99-104.

Turner, D. [2007]: *Making Prehistory – Historical Science and the Scientific Realism Debate.* Cambridge: Cambridge University Press.