

Delimiting the Unconceived

Richard Dawid*

December 20, 2017

Abstract

It has been argued in Dawid (2013) that physicists at times generate substantial trust in an empirically unconfirmed theory based on observations that lie beyond the theory's intended domain. A crucial role in the reconstruction of this argument of "non-empirical confirmation" is played by limitations to scientific underdetermination. The present paper discusses the question as to how generic the role of limitations to scientific underdetermination really is. It is argued that assessing such limitations is essential for generating trust in any theory's predictions, be it empirically confirmed or not. The emerging view suggests that empirical and non-empirical confirmation are more closely related to each other than one may expect at first glance.

*Department of Philosophy, Stockholm University, 10691 Stockholm (Sweden) – richard.dawid@philosophy.su.se

1 Introduction

Fundamental physics today is characterized by a scarcity of empirical data that leaves pivotal theories without conclusive empirical testing (sometimes without any empirical testing at all) for many decades. In order to assess the status of those theories, physicists resort to various strategies of theory assessment. In some cases like string theory or eternal inflation, those strategies, in the understanding of many of the given theories' exponents, justify a fairly high degree of trust in the theory's viability. Dawid (2006, 2013) and Dawid, Hartmann and Sprenger (2015) have analyzed the question whether arguments of non-empirical theory assessment can be epistemically significant. It was claimed that a specific class of those arguments can be reconstructed in a way that demonstrates their epistemic significance and, from a Bayesian perspective, justifies calling them "non-empirical confirmation". In conjunction and under the right circumstances those arguments can lead to a substantial increase of the theory's subjective probability of being viable.

The instantiations of non-empirical confirmation presented in Dawid (2013) share one crucial conceptual element: they are based on the assessment of limitations to local scientific underdetermination. Scientific underdetermination measures how many alternative theories that are not empirically fully equivalent to each other can account for a given empirical data set. Local scientific underdetermination accounts only for those theories that could be distinguished within a given "experimental horizon" specified by the reach of a class of possible experiments. According to Dawid (2013), substantial trust in a theory in the absence of empirical confirmation can be generated based on a certain kind of meta-level observation about the research process. Those observations support the hypothesis that only very few if any scientific alternatives to the theory at hand exist that are empirically distinguishable within a given empirical horizon. Non-empirical confirmation is based on observations of that kind.

An important question that arises in this context is the following: how generic is the role of limitations to scientific underdetermination when it comes to significant arguments of non-empirical theory assessment? The present paper addresses this question by focusing on the case of empirical confirmation. It concludes that limitations to underdetermination not only are an essential element of non-empirical theory confirmation. They play a fundamental role in understanding empirical confirmation as well. Assessments of limitations to scientific underdetermination therefore do not appear as a whimsical way out for scientists trying to believe in their theory's value in the absence on empirical testing. Quite to the contrary, they constitute a

core element of the scientific process whenever confirmation, be it empirical or non-empirical, is involved. Empirical and non-empirical confirmation in this light are much closer connected to each other than it would seem at first glance.

After briefly introducing the three arguments of non-empirical theory confirmation in Section 2, Section 3 argues for a close connection between theory confirmation and trust in a theory's predictions. Following an aside on truth and viability in Section 4, Sections 5 and 6 demonstrate the importance of scientific underdetermination for developing a satisfactory understanding of empirical confirmation. Section 7 discusses repercussions of the presented point of view for one line of criticism against non-empirical confirmation.

2 The Three Arguments of Non-Empirical Theory Confirmation

Dawid (2013) proposes three main arguments of non-empirical confirmation.

NAA: The No Alternatives Argument: 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 is taken as an indication of the viability of theory H .

MIA: The Meta-Inductive Argument from success in the research field: Theories in the research field that satisfy a given set of conditions have shown a tendency of being viable in the past. This observation is taken to increase the probability that a new theory H that also satisfies those conditions is also viable.

UEA: The Argument of Unexpected Explanatory Interconnections: Theory H was developed in order to solve a specific problem. Once H was developed, physicists found that H also provides explanations with respect to a range of problems which to solve was not the initial aim of developing the theory. This observation is taken as an indication of the theory's viability.

Each of the three arguments remains weak and questionable in isolation. But the arguments gain strength and significance in conjunction. They play an important though often contested role in generating trust in empirically unconfirmed or insufficiently confirmed theories.

3 The role of scientific underdetermination

One might feel content with having identified the three arguments of non-empirical theory assessment in the physicists' reasoning and leave it to the specifics of the physical debate to clarify how satisfactory those arguments are in a given context. This strategy, however, seems insufficient for answering a general but important question. What is the status that can in principle be acquired by arguments of non-empirical theory assessment? Do they carry substantial epistemic value? And to the extent they do, how can their epistemic value be understood within the general fabric of scientific reasoning? In order to get a grip on this question, it seems important to understand the arguments' general conceptual foundations.

Analysing the issue at a basic level starts with reconsidering the general role of evidence and confirmation in science. The canonical view on the scientific process insists that epistemically significant support for a theory's viability can only be attained based on empirical evidence. On that view, non-empirical evidence would lack epistemic significance and be confined to the purely pragmatic role of influencing strategic decisions regarding the selection of the theory one intends to work on in the absence of empirical guidance.

As argued in more detail in Dawid (forthcoming), this understanding is unsatisfactory for two reasons. First, it is at variance with the degree of trust many physicists do have in empirically unconfirmed theories in the absence of empirical confirmation. Second, it seems somewhat inconsistent to admit that arguments of non-empirical theory assessment play a substantial and legitimate role in a scientist's selection of the theory she wants to work on without conceding that this substantial role is rooted in epistemologically significant analysis. After all, the eventual goal of a scientific theory is to be physically viable. Considerations which do not increase the subjective probability of a theory's viability thus offer only limited help for making strategic decisions. In this light, it is important to understand whether non-empirical theory assessment can be construed in a way that reaches out beyond the canonical understanding of epistemic significance in the scientific process.

Dawid (2013) makes the suggestion that all three arguments can be reconstructed in terms of arguments for limitations to local scientific underdetermination and get epistemic significance on those grounds. It is argued that there exists a positive probabilistic correlation between each of the three arguments and the hypothesis that only very few possible theories - or maybe just one theory - exist that can account for the observed physical situation. This is due to the fact that a hypothesis stating strong limitations to local

scientific underdetermination can explain the observations spelled out in the three arguments of non-empirical confirmation

NAA: The fact that scientists don't find alternatives to a theory at hand can be explained by the hypothesis that there are few or no possible alternatives to that theory.

MIA: The fact that theories in a research field that satisfy a given set of conditions have high chances of predictive success can be explained by the hypothesis that there tend to be few alternatives in the given research field under the given conditions.

UEA: The fact that unexpected explanatory interconnections are found can be explained by the hypothesis that there are fewer alternative theories in the given context than there are conceptual aspects of the phenomenology that await explanation.

A scarcity of possible theories, in turn, increases the probability that the theory that has been found is viable. Therefore, the three arguments on non-empirical confirmation establish a higher probability of the theory's viability.

Is this reconstruction of non-empirical confirmation in terms of limitations to scientific underdetermination an arbitrary choice? The present paper aims to demonstrate that, quite to the contrary, non-empirical confirmation is just a particularly conspicuous case of the high importance of assessments of limitations to scientific underdetermination in the scientific process.

I want to start the analysis with a pair of questions on empirical confirmation. Can confirmation of a fundamental theory by empirical data be the basis for trusting that theory in future research contexts? And to the extent it does, how can this connection be understood? Philosophy of science has often been very timid in attempts to answer this question. Karl Popper has famously claimed that science does not confirm theories but merely tests them with the aim of refuting them. On his account, science is not in the business at all of assessing a theory's chances of future correct predictions.

Bayesianism¹, the leading theory of confirmation today, chooses a different perspective. Bayesians understand confirmation in terms of the increase of a theory's truth probability under observational evidence and emphasize that confirmation does play a central role in scientific reasoning. However,

¹for surveys of Bayesian epistemology, see e.g. Bovens and Hartmann (2003), Howson and Urbach (2006), Strevens (2017)

neither the precise meaning of the statement that a theory is true nor the numerical values of truth probabilities extracted based on Bayesian updating take center stage in Bayesian epistemology.

The main reason for this restraint lies in the philosophical difficulties faced by the concept of truth. Scientific data analysis measures a theory's performance in a given experimental setting. It may aim at rejecting a null hypothesis based on collected data or at comparing rivaling theories based on collected data. With respect to those issues, convergence theorems tell us that in the limit of infinitely many data points, in a large class of scientific contexts the results of Bayesian analysis will converge towards frequentist results. The effect of subjective priors thus can be "overpowered" by data. A sequence of empirical testing can establish results that are eventually endorsed by the entire scientific community irrespectively of the diversity of the scientists' prior expectations.

What data analysis cannot do is quantify a scientific theory's chances of being absolutely true. Any attempt to do so would require a forecast with respect to the theory's performance under all possible future empirical tests. Data analysis with respect to one empirical test offers no scientific basis for a forecast of that kind.

The issue thus must be decided at a philosophical level within the framework of the scientific realism debate, addressing issues from the pessimistic meta induction to the question as to whether or not empirical adequacy implies truth. Differences in priors that are due to different positions regarding those philosophical questions cannot be expected to vanish with increased data volume since the philosophical considerations involved may not be thoroughly testable by empirical data at all.

The stated problems for specifying absolute truth probabilities cannot be avoided by resorting to objective Bayesianism. It is questionable whether an objective Bayesian would want to extend the reach of objective priors all the way to settling the scientific realism debate. Even if she assumed that objective priors on scientific realism exist, she might refrain from asserting that she is able to specify them at the present point. And even if she did, she might prefer characterizing in Bayesian terms the scientists' understanding of a given theory's current epistemic status without imposing on them her own preconceptions regarding philosophically highly charged priors.

For the reasons just stated, Bayesian epistemologists, be they objective or subjective Bayesians, refrain from discussing absolute truth probabilities of scientific theories. The effectiveness of Bayesian confirmation theory stems from the fact that the attribution of Bayesian confirmation is invariant under the variation of priors (as long as dogmatic priors 1 and 0 are excluded). The Bayesian differential definition of confirmation as a probability increase

provides the basis for analysing a wide range of issues on confirmation without ever specifying actual probabilities.

Still, this strategy carries far-reaching implications. Though the Bayesian perspective implicitly suggests a correlation between theory confirmation and the degree of trust one is ready to invest in a theory's future performance, the approach's restraint regarding the absolute specification of truth probabilities keeps it from spelling out the quantitative specifics of the former correlation as well. The Bayesian thus offers an account of confirmation that refrains from answering the question how much trust a scientist should in fact have in a specified class of predictions of a confirmed theory.

Popperian falsificationalism and canonical forms of Bayesian confirmation theory, for all their differences, thus share one problem. They don't account for the fact that, for the scientist, theory confirmation is closely linked to the question whether she can trust the confirmed theory's new predictions.

To understand how crucial that issue becomes in actual science, let us briefly look at an example from recent high energy physics. In Summer 2012, the ATLAS and CMS experiments at the LHC at CERN both announced an effect with significance above five sigma indicating the existence of a new scalar particle that, after further detailed analysis, was firmly established as a Higgs-like particle: a particle that in crucial respects shares the characteristics of the Higgs field that had been posited to explain the mass spectrum of elementary particles in the standard model.

The Higgs discovery was based on very specific kinds of signatures that showed up in ATLAS and CMS. The implications of the existence of a Higgs field, however, go beyond the specific role played by the Higgs field in explaining the frequency of specific vertices in those two experiments at the LHC. For example, quantum field theory predicts that a Higgs field, apart from resulting in the generation and decay of actual particles in scattering events, also contributes via off mass shell effects in situations where there is not sufficient energy available for generating an actual Higgs particle. This process of virtual particle exchange contributes to scattering processes and has an effect on the predictions of scattering amplitudes. Announcing the discovery of a particle amounts to the announcement that all physical effects of that particle, including the described contributions of virtual particles, must be included in all future calculations of scattering amplitudes. Doing this right is essential for being able to make new discoveries of particles which, like in the case of the Higgs discovery, must be based on comparing measured event rates that may contain the new particle with the calculated background (i.e. the expected event rate in the absence of that new particle). Only if the calculations of the background are reliable, the search for new particles can proceed in an effective way.

Scientific progress in the field therefore crucially depends on the physicists' justified trust in the entire set of implications of an empirically confirmed theory like the Higgs theory. In sharp contrast to Popper's ideas, science is not only about refutation. It is also and importantly about trusting well-confirmed theories. A full understanding of the scientific process therefore hinges on understanding how scientists can develop that trust. Any attempt to do so, however, must be based on absolute probabilities that can characterize a theory's status.

In order to understand the difficulties associated with finding such a probabilistic construal of theory assessment in fundamental science, it is helpful to remember that everyday reasoning avails itself of a seemingly unproblematic strategy of generating trust in a somewhat similar context. Let us consider a philosophically well known example of a hypothesis in the context of everyday reasoning. I hear a scratching sound behind the wainscoting and note that breadcrumbs dropped on the floor in the evening have disappeared in the morning. I infer from this set of observations that there probably is a mouse behind the wainscoting.

Clearly, a wide range of possible explanations of the scratching sound exist. The more inventive I become in developing them, the more the set of explanations to be considered will grow. It does not make much sense to even try to keep track of all possible explanations. Nevertheless, I may feel fairly confident that my mouse hypothesis is correct.

The reason for my confidence lies in my high degree of confidence in my overall theory of my environment. I believe that this overall theory, while being consistent with a wide range of explanations of the wainscoting observation, allows me to assess the probability of each of those possible explanations, including the catch-all hypothesis that covers everything I could not even think of. I have access to a long record of household events that tell me that en-passant observations in apartments have very rarely led to fundamental new discoveries about the world. On that basis, I will attribute a very low probability to that possibility. I am also fairly confident to have sufficient knowledge about the known possible explanations. So I scan through the options I can think of. I know that goblins aren't real. I take the probability of an extraterrestrial attack on my apartment to be negligible. I think that the other human inhabitants of my apartment are not technically sufficiently versed to play a trick on me by generating a genuinely mouse-like sound behind the wainscoting. But the mouse possibility is very plausible and I had mice in the apartment before. Based on all of these complex steps of analysis which are grounded in my overall theory about the world and my neighborhood, a probabilistic analysis leads me to endorse the mouse hypothesis with considerable confidence.

The described line of reasoning is a typical example of inference to the best explanation in an everyday context. It is based on carrying out an exhaustive probabilistic assessment of all explanations that seem possible based on my well established and well tested world view. Being philosophically informed, I'm aware of the fact that the problem of induction lurks somewhere in the background. But it seems of limited practical relevance.

Unfortunately, the scheme just presented only works as long as reasoning happens within the safe confines of the well known world. It is an essential precondition for the success of my wainscoting analysis that I can attribute a low probability to the possibility that the core principles of my theory about the world are inconsistent with the true explanation and have to be modified to account for my observation. This may be taken for granted in most everyday life situations. But it cannot be assumed in fundamental science.

In fundamental science, one is typically confronted with data that indeed is at variance with our well established fundamental theory. Scientists search for a new theory that can account for the new data. The question that corresponds to the earlier question as to how much confidence I could have in the mouse hypothesis without having seen the mouse now becomes: how much trust can scientists have in the future predictive success of a hypothesis that was developed in order to account for a given set of anomalous data? The kinds of reasoning that led to a probabilistic appraisal of possible explanations in the wainscoting example is of little use here. Since the observations to be accounted for are known to be in conflict with our well-established theories, it makes no sense to assess the probability of a suggested new theory within the old theoretical framework. If scientists find a novel theory T_a in fundamental science that accounts for some anomalous data E_1 , considerations on the trust one may have in T_a and its so far untested empirical predictions thus boil down to assessing precisely the one aspect of the wainscoting analysis that had been considered safe to disregard in that context: we need to assess the probability that the data E_2 to be collected at the next step of empirical testing will be consistent with a different fundamental theory T_b rather than with T_a .

I have argued above that scientists do generate such trust and must rely on it in order to do meaningful science. I have also argued that canonical theories on theory testing and confirmation don't offer good reasons for this trust. So how could we find such good reasons? The answer will lead back to limitations to scientific underdetermination. Section 5 will make the case that the mechanism that generates trust in the predictions of empirically confirmed theories is of exactly the same type as the mechanism that can support non-empirical confirmation in the absence of empirical confirmation.

4 What Scientists mean when they endorse a Theory

As pointed out above, physics relies on an implicit quantitative assessment of the degree of trust one should have in a theory. But the trust physicists are aiming at has little to do with the issue of a theory's truth. When they endorse a theory, their endorsement is perfectly consistent with their expectation that the theory has a limited range of applicability and will have to be superseded by a more fundamental theory once one aims at describing a wider range of empirical data. To give one example, endorsing the standard model of particle physics as an adequate description of physics up to the electroweak scale does not imply the rejection of theories positing new physics at higher energy scales. Those more fundamental theories may well be based on a substantially different ontology than the standard model and therefore be at variance with the truth or even, in an ontological sense, approximate truth of the latter. The standard model's endorsement by physicists thus remains entirely independent from philosophical considerations about the theory's truth.

Truth in this light does not appear to be a helpful concept for understanding theory assessment by scientists. I suggest that the most effective way of representing the scientist's perspective on theory confirmation consists in understanding theory assessment and confirmation in terms of what I call a theory's viability rather than its truth. The idea is

- 1: to rely on a Bayesian approach in order to account for the probabilistic nature of theory assessment.
- 2: to reject truth as the object of probabilistic analysis in this context.
- 3: to introduce the concept of empirical viability within a given empirical context as the basis of probabilistic analysis. This step allows to link scientific confirmation to the ascription of absolute probabilities.

On this account, confirmation is understood as an observation-based increase of the probability that the confirmed theory is viable within a given empirical horizon. An example of specifying an empirical horizon in the context of high energy physics would be specifying an energy scale up to which a theory is tested. Viability is then defined as the agreement of the theory's predictions with all empirical data that can be possibly collected within that

empirical horizon. The probability of a theory's viability therefore can be specified only with respect to a spelled out empirical horizon. Henceforth, we will write

$$P(T) \tag{1}$$

as the probability that theory H is viable. Strictly speaking, T should carry an index denoting the chosen empirical horizon. We will omit that index for the sake of simplicity.

The claim that the three lines of reasoning presented in Section 2 can in conjunction amount to significant confirmation can now be understood in this framework: the three arguments can substantially increase the subjective probability that a theory be viable within a given empirical horizon.

5 A spectrum of Choices

Scientific theories make predictions about observable phenomena. Empirical data that can be in agreement or at variance with a theory's predictions are said to be within the theory's intended domain. Bayesian confirmation is canonically understood to rely on empirical data within a given theory's intended domain. The idea that observations that lie beyond a theory's intended domain can amount to Bayesian confirmation has not been considered in canonical Bayesian epistemology for two reasons. First, as will be discussed in a little more detail later on, the focus on comparing a theory's predictions with empirical data offers the most straightforward way of delimiting the scientific method from other modes of reasoning. Second, the direct argument for an increase of a theory's probability due to an observation is not applicable to observations that lie beyond the theory's intended domain.

For data E that lies within the intended domain of theory H , $P(E|T)$ can be extracted from H itself. Bayes' formula

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

implies that data E confirms H if $P(E|T) > P(E)$. A moderate value for $P(E)$ indicates the a priori understanding that the viable theory describing the given physical context might well be one that does not predict E . A value of $P(E|T)$ that lies substantially above what seems a plausible value of $P(E)$ then amounts to substantial confirmation.

With respect to observations F that lie outside the theory's intended domain, $P(F|T)$ cannot be extracted from the theory H itself. Therefore, no immediate line of reasoning leads from the theory's empirical implications

towards establishing confirmation. This does not rule out, however, that non-empirical evidence F may have confirmation value with respect to H . The most straightforward way of constructing a scenario where this is the case is to retain the mechanism of confirmation based on a prediction at a meta-level. The data F is then taken to be predicted by some meta-level hypothesis Y . In other words, while F lies outside the intended domain of theory H , it lies within the intended domain of the meta-level hypothesis Y . The crucial question then becomes whether there is a positive probabilistic correlation between the truth of Y and the viability of H . If that is the case, F confirms H via confirming Y .

As described in Section 3, limitations to scientific underdetermination can step in at this point. The hypothesis Y that can be used most effectively for establishing a confirmation value of all three arguments of non-empirical confirmation is a hypothesis on strong limitations to scientific underdetermination. In the following, I want to argue for the cogency of this suggestion within a Bayesian framework by embedding it in the general view on confirmation developed above. It will turn out that, from that perspective, the significance of limitations to scientific underdetermination can be understood without thinking about non-empirical confirmation at all.

$P(E)$ in (2) can be written as a total probability

$$P(E) = \sum_{i=1}^n P(T_i)P(E|T_i) + P(CA)P(E|CA) \quad (3)$$

for n known theories H_i , T_i being the statement that theory H_i is viable. CA denotes the catchall hypothesis that an unconceived alternative theory rather than any of the known alternatives is viable. (Remember that we have specified theory individuation based on distinguishable predictions within a given empirical horizon.) Let us consider the case where one known theory H correctly reproduces a set of data E_1 . No alternatives to H are known that reproduce E_1 . The question is how much trust scientists should have in further predictions of H within an empirical horizon E_2 . CA covers all unconceived alternatives that correctly represent E_1 . Trust in the scientific process suggests to attribute a very low probability to the possibility that the viable theory in the given context is a theory that violates core principles of scientificity. This includes disregard for theories that are manifestly inconsistent in a way that prevents a cogent set of physical implications, for theories that are predictively empty, for theories that fail in a substantial way to reproduce the predictive power of empirically well confirmed theories that are understood to work as the new theory's effective theory and probably some other categories of deeply unsatisfactory theories. Therefore, we expect that

$$P(CA) \simeq \sum_{j=1}^m P(T_j)P(E|T_j) \quad (4)$$

summing over the m unconceived alternatives that satisfy core conditions of scientificity. In the absence of further knowledge about the characteristics of those unconceived scientific alternatives, it seems very plausible to make two general assumptions.

A: each of the unconceived alternatives has roughly the same prior probability of being viable.

B: an unconceived alternative has roughly the same prior probability of being viable as the known candidate theory H .

Both assumptions can be motivated by considering theory assessment with respect to known alternatives. Assumption A looks plausible in light of the way a spectrum of known alternatives is understood to affect the prospects of an individual theory. If a spectrum of known alternatives is known to account for the data, scientists are ready to attribute roughly equal probabilities to the viability of each of them as long as they have no strong reason to favor or disfavor any of them. This indeed leads to a direct dependence of the trust they are willing to invest into an individual theory on the number possible alternatives they are aware of. For example, the vast number of models of supersymmetric theories, despite the fact that they differ in simplicity and with respect to their conceptual merits, implies that no individual model is taken to be trustworthy, even under the assumption that low energy supersymmetry applies at all. The fact that considerations along those lines are used with respect to a spectrum of known theories where differences with regard to the individual theories' merits can be spelled out makes it even more plausible to deploy this kind of consideration in the case of unconceived alternatives where no such differences can be specified.

Assumption B can be argued for based on the way a newly developed alternative theory is treated. If a new theory with comparable scientific merits is developed as a possible alternative to a long known theory, the new theory is taken to have prospects of being viable that are comparable to those of the earlier theory. The fact that the new theory had not been discovered earlier is not taken to be a significant argument against that theory's viability. For example, when large extra dimensions plus brane physics were understood to provide an alternative scenario for understanding the grand unification scale, that scenario was considered a serious alternative with significant prospects of viability irrespectively of the time line of theory development. Given that

the merits of unconceived alternatives in comparison with the known theory's scientific merits cannot be specified, it seems natural in this light to assume prior probabilities of an unconceived alternative's viability that are comparable to those attributed to the known theory.

In conjunction, assumptions A and B amount to the understanding that the theory that has been developed can roughly be viewed as a random pick out of the ensemble of possible alternatives. It is assumed that scientists don't own a truth detector that guides them towards discovering the viable theory first. The described view implies that the prior probability attributed to the catch all hypothesis is directly linked to the assessment as to how many unconceived alternatives to a given empirically confirmed theory exist that predict the confirming data as well. As long as that number is taken to be large, the catch-all hypothesis dominates the probability assessment and reduces the probability of the known theory's viability within the empirical horizon E to very small values. There is no basis for trusting the predictions of theory H . Strict limitations to scientific underdetermination are therefore necessary for having trust in a theory's empirical predictions.

6 The Wider Picture

Let us put the above Bayesian analysis of the connection between theory confirmation and trust in a theory's predictions into the wider context discussed in Section 3.

The plausibility of extending the concept of confirmation to include observations beyond the confirmed theory's intended domain has been overlooked or under-appreciated in the philosophy of science up to this point due to two deeply entrenched ways of thinking about the scientific process.

i: Since confirmation has traditionally been linked to truth, the philosophical problems related to the concept of truth have led to a decoupling of the analysis of confirmation from the issue of extracting absolute probabilities of a theory's viability in a given empirical context. This has resulted in a split between the issue of confirmation and the question as to how scientists can actually justify their trust in further predictions of an empirically confirmed theory.

ii: The distinction between scientific and non-scientific reasoning has largely been built on the testability of scientific theories by empirical data within the theory's intended domain. Scientific theory confirmation, to the extent it has been acknowledged as a core element of scientific reasoning, has mainly been understood as the process that implements empirical testing as a cornerstone of scientific reasoning. Therefore it has been taken to be

based, virtually by definition, on empirical data within the theory's intended domain.

However, divorcing the notion of confirmation from the notion of specifying actual trust in a theory's predictions leaves out an essential element of the role played by confirmation in science: the importance of confirmation to the scientist rests to a considerable degree on the trust it licenses in a theory's predictions. This implies that an adequate analysis of the role of confirmation in science needs to address the argumentative mechanisms that lead from confirming empirical data to trust in a theory's predictions.

Once the issue of trust in a theory's predictions has been clearly separated from the issue of truth, it becomes clear that the former hinges on the the understanding that local scientific underdetermination is very limited. But limitations to local scientific underdetermination cannot be inferred from data within the theory's intended domain. They need to be inferred based on arguments of non-empirical theory assessment.

Empirical confirmation thus emerges as a compound of two levels of analysis. Confrontation of a theory's predictions with empirical data establishes that a theory is empirically confirmed. On that basis, non-empirical theory assessment establishes claims of limitations to scientific underdetermination that determine the confirmation value of the confirming data in terms of the degree to which it licenses trust in the theory's further empirical predictions. The strength of limitations to scientific underdetermination crucially depends on the strength of the confirming data: the more striking the agreement between the collected data and a theory's predictions, the fewer alternative theories may be expected to match that predictive success. However, even the strongest agreement of a theory's predictions with collected data does not license trust in the theory's further predictions in the absence of an assessment of limitations to scientific underdetermination.

Empirical testing remains the cornerstone of confirmation. Empirical viability is defined via empirical testing and empirical testing someplace else in the research field constitutes one crucial element of non-empirical theory assessment (MIA). In one respect, however, the presented view actually inverts the hierarchy between empirical testing and non-empirical theory assessment. Non-empirical theory assessment of the theory under scrutiny emerges as the only irreplaceable element of analysis that is necessary for generating significant trust in that theory's predictions. In the absence of arguments indicating that scientific underdetermination is limited in the given context, even the strongest agreement between theory and data would not lead to trust in the theory's further predictions. The probability of the theory's viability would not get significantly increased.

The agreement between collected data and a theory's quantitative core

predictions offers the by far most powerful basis for extracting claims of limitations to scientific underdetermination. But it is not the only kind of information that can provide a basis for considerations on limitations to scientific underdetermination. Limitations to scientific underdetermination may also be considered with respect to general characteristics of the collected data that stand in a highly indirect relation to the core characteristics of the theory that has been developed to explain them. The Higgs mechanism has long been suspected to be the only scientific conceptual scheme that could explain the mass spectrum of elementary particles while retaining the explanatory power of gauge theory in the given context. This conjecture of strong limitations to scientific underdetermination was put forward long before empirical signatures of Higgs particles had been found. String theory is taken by many of its exponents to be the only scientific theory that can explain the observed structure of nuclear interactions that can be represented by a gauge theory plus the existence of a gravitational force. This claim of strong limitations to scientific underdetermination is made in the absence of any empirical data that can be understood as the signature of a string.

In cases like these, trust in the theory's predictions is generated already at a stage where none of the theory's predictions has been tested - or even quantitatively spelled out. Non-empirical theory assessment is then deployed as actual non-empirical theory confirmation. In our world, non-empirical confirmation will never reach the levels of conclusiveness of strong empirical confirmation. But the difference between the two kinds of confirmation is gradual. Non-empirical theory assessment plays a crucial role in both of them and reveals a strong continuity between empirical and non-empirical confirmation.

7 The reliability of non-empirical theory assessment

The above reasoning offers a fresh perspective on one important line of criticism that has been put forward against non-empirical confirmation. Crucial elements of the assessment of limitations to scientific underdetermination, such as the specification of scientificity criteria that have to be fulfilled by a scientific alternative or the specifics of theory individuation are kept fairly vague. Dardashti (forthcoming) and Oriti (forthcoming) have argued that these vaguenesses render statements of non-empirical confirmation underspecified and therefore threaten their significance.

The present text is not concerned with the structural analysis of this issue.

However, the analysis presented in the previous sections can put the status of the described worry into a wider perspective. We have seen that assessments of limitations to scientific underdetermination are of crucial importance for building up trust in an empirically confirmed theory's further predictions. An adequate representation of the way scientists understand the significance of empirical confirmation must therefore acknowledge assessments of limitations to scientific underdetermination as an integral element of the process of theory confirmation. The very same vaguenesses, however, that characterize arguments for limitations to scientific underdetermination in the context of non-empirical confirmation also arise when those arguments are deployed for understanding the predictive reliability of an empirically confirmed theory.

One might bite the bullet and argue that empirical confirmation in the wider sense I suggest fails to be a workable concept as well. This step, however, would lead back to the unsatisfactory situation described in Section 3. It would ignore the fact that scientists do link confirmation to appraisals of the trustworthiness of a theory's predictions. And, even more significantly, it would ignore the fact that scientists have a very high degree of trust in further predictions of empirically well confirmed theories (remember the example of the strong reliance on contributions of newly discovered particles to the calculation of the background for scattering amplitudes in high energy physics).

The very stable role of empirically well confirmed theories in the research process demonstrates that the vagueness of some elements of non-empirical theory assessment does not per se rule out very high posteriors for a theory's viability. This does not mean that the described vaguenesses are conceptually unproblematic. It indicates, however, that the problems that arise when aiming at a precise understanding of statements of limitations to scientific underdetermination should not be taken as an argument against those statements' epistemic relevance. A careful conceptual analysis of those vaguenesses is just as important for understanding empirical confirmation as it is for understanding non-empirical confirmation.

8 Conclusion

Theory confirmation may be approached in two different ways. One may either understand confirmation formally in terms of an increase of a theory's truth probability. Or one may understand confirmation in terms of the actual degree of trust generated with respect to a theory's predictions. The first view has the advantage of avoiding the tricky issue how to translate outcomes of past experiments into prospects for the outcome of new ones. The second

view, on the other hand, has the merit of addressing the main reason why confirmation is of crucial importance in science.

If the second view on confirmation is chosen, it seems difficult to avoid the conclusion that assessments of limitations to local scientific underdetermination play an important role in specifying the confirmation value of confirming data. Empirical confirmation then relies both on the confirming data and on non-empirical theory assessment that establishes claims of limitations to local scientific underdetermination. Only in conjunction, the two elements of empirical confirmation can generate substantial trust in the theory's further predictions.

Once this has been acknowledged, it is a small step towards conceding that confirmation may also be realized in the absence of empirical testing of a theory if other considerations provide a framework for inferring limitations to scientific underdetermination. Non-empirical and empirical confirmation in this light are conceptually related via the concept of limitations to scientific underdetermination.

References

- Bovens, L., and Hartmann, S. (2003): *Bayesian Epistemology*, Oxford: Oxford University Press.
- Dardashti, R., (forthcoming): "Science Without Experiments", to appear in *Why Trust a Theory? - Reconsidering Scientific Methodology in Light of Modern Physics*, Dardashti, R., R.Dawid and K. Thebault (eds.), Cambridge University Press.
- Dawid, R. (2006): "Underdetermination and Theory Succession from the Perspective of String Theory", *Philosophy of Science* 73/3, 298-322.
- Dawid, R. (2013): *String Theory and the Scientific Method*, Cambridge: Cambridge University Press.
- Dawid, R. (forthcoming): "The Significance of Non-Empirical Confirmation in Fundamental Physics", to appear in *Why Trust a Theory? - Reconsidering Scientific Methodology in Light of Modern Physics*, Dardashti, R., R.Dawid and K. Thebault (eds.), Cambridge University Press.
- Dawid, R. S. Hartmann and Jan Sprenger (2015): "The No Alternatives Argument", *The 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.
- Oriti, D. (forthcoming): "No Alternative to Proliferation", to appear in *Why Trust a Theory? - Reconsidering Scientific Methodology in Light of Modern Physics*, Dardashti, R., R.Dawid and K. Thebault (eds.), Cambridge University Press.
- Strevens, M. (2017): "Notes on Bayesian Confirmation Theory", online lecture notes.