Novel Confirmation and the Underdetermination of Scientific Theory Building

Richard Dawid

University of Vienna,
Department for Philosophy and Institute Vienna Circle
Universitätsstraße 7. 1010 Vienna, Austria
email: Richard.dawid@univie.ac.at
tel: +43 650 7287299

The extra value of novel confirmation over accommodation is explained based on an analysis of the underdetermination of scientific theory building. Novel confirmation can provide information on the number of possible scientific alternatives to a predictively successful theory. This information, in turn, can raise the probability that the given theory is true or will be empirically viable in the future.

Keywords: novel confirmation, predictivism, Bayesianism, underdetermination,

1: Introduction

The debate on novel confirmation addresses one of the classical questions in philosophy of science. Does confirmation by novel empirical data (novel confirmation) provide stronger confirmation of a theory than the consistency of a theory with empirical data that had entered its construction process (accommodation)? The predictivist position holds that it does, which seems to agree with what a majority of scientists takes to be the intuitive understanding of scientific reasoning. Many philosophers of science throughout the centuries have rejected predictivism, however. In the 19th century, Whewell endorsed predictivism while Mill rejected it. In the mid 20th century, Popper and Lakatos supported the significance of novel prediction in their own ways. Since the 1990s, the debate has gained new impetus within a Bayesian framework. While Bayesianism was initially held by many to be incompatible with predictivism, it was soon understood that a sufficiently general Bayesian reconstruction can account for a confirmation extra value of novel confirmation (Maher 1988, Kahn et al. 1994, Barnes 2008). Still, a number of influential philosophers (see e.g. Worrall 2002, Mayo 1996) have argued in recent years that an emphasis on novel confirmation fails to grasp the core mechanisms of confirmation.

In this paper, I present a new argument for predictivism. This argument is based on the concept of 'limitations to scientific underdetermination' that has been discussed in a number of recent works (Dawid 2006, Dawid 2010, Dawid 2013, Dawid, Hartmann and Sprenger forthcoming,) in the context of empirically unconfirmed theories. According to those works, the assessment of the number of possible alternatives to the theory scientists have actually developed can shed light on that theory's success chances even before the theory has found empirical confirmation. Therefore, it is argued, such 'assessments of scientific underdetermination' can themselves constitute theory confirmation under certain conditions. In the following, this approach shall be applied to the question of novel confirmation. It shall be argued that novel confirmation, unlike accommodation, tells something about the number of possible alternative scientific theories and thereby has confirmation extra value over accommodation.

The presented approach shares some general characteristics with other suggestions that have been put forward for explaining novel confirmation in recent decades (Maher 1988, Kahn et al. 1994, Hitchcock and Sober 2004, Barnes 2008). It does not aim at fully replacing those earlier suggestions. Various individual reasons for an extra-value of novel confirmation can apply in science, and each of them may be of particular significance in certain scientific contexts. The presented approach adds to that list of reasons and, as will turn out, is of particular relevance in fundamental physics. I will argue, however, that the significance of the presented approach goes beyond merely providing an additional aspect of novel confirmation. As I hope to demonstrate, the significance of assessments of scientific underdetermination for the question of novel confirmation integrates the latter into a broader understanding of the mechanism of theory confirmation in science. It establishes that the extra value of novel confirmation is by no means tangential to the core mechanisms of scientific reasoning but exemplifies a crucial element of the way scientists determine and understand the status of their theories. After a brief introduction to the novel confirmation debate in Section 2, the core argument is presented in Sections 3 and 4. Section 5 compares the presented approach to other suggestions in the literature. Section 6 then looks specifically at the case of fundamental physics. Some conclusions are presented in Section 7.

2: The Problem of Novel Confirmation

Following Zahar (1973) and Worrall (1978), I take novel confirmation to be confirmation of a theory H by data that did not play a heuristic role in the development of the theory. I further require that novel confirmation is constituted by data that is not just 'more of the same' but amounts to a new phenomenon set apart from what has been observed before. The question now is whether or not novel confirmation in the given sense has higher confirmation value than accommodation, i. e. a theory's agreement with empirical data that did play a heuristic role in its development.

The problem of novel confirmation can be formulated in a transparent way within a Bayesian framework. In order to see the problem in this context, let us first confine the discussion to empirical data within the intended domain of the scientific theory to be analysed. That is, we only consider data of the kind that might in principle be predicted by the scientific theory H. Let us assume that some data E is collected in an experiment. The Bayesian formalization then simply expresses the relation between the probability that a theory H is true if evidence E were the case and the probability of the truth of H when E is not taken into account. If the data E is predicted by theory H, the truth of E raises the

probability of the truth of H. E therefore confirms H. Contingent historical facts regarding the influence data E had on the construction process of H are not expressible in terms of data that is predicted by H. Therefore, such facts do not enter the analysis. Novel confirmation on that basis cannot be awarded higher significance than accommodation.

It is generally acknowledged today that the above analysis on its own does not refute predictivism. The Bayesian approach can be extended to include empirical data that is not predictable by the theory under investigation. In particular, one can take into account the observation that some empirical data E has not entered the construction process of theory H. While this observation is of an empirical nature, theory H does not predict whether or not it occurs. Once one includes data of the described kind into a Bayesian analysis, novel confirmation can get additional confirmation value compared to accommodation. Let $O^H(E)$ be the observation that E has not entered the construction of H and $\not{O}^H(E)$ the observation that E has entered the construction of H. It may then be the case that

$$P\left(T\middle|E \wedge O^H(E)\right) > P\left(T\middle|E \wedge \not{O}^H(\mathrm{E})\right) \qquad (1)$$

where T is the statement that theory H is true. Justifying statement (1) requires additional conceptual input, however. Since neither $O^H(E)$ nor $\not{O}^H(E)$ constitute evidence of a kind predictable by theory H, it is not immediately clear on what grounds that kind of evidence can change the probability of the truth of theory H. The most straightforward way to establish that it does is to find a new class of statements Y that can predict occurrences of the kind $(E \wedge O^H(E))$. Therefore, $(E \wedge O^H(E))$ will under normal circumstances confirm Y. If the truth of Y, in turn, raises the probability of the truth of H, the extra-value of novel confirmation is established.

The described line of reasoning was followed first in Maher (1988) and later by a number of other authors. These authors make various suggestions regarding the nature of the claim that is directly supported by $(E \wedge O^H(E))$. Maher proposes that novel confirmation tells something about the quality of the method of scientific theory construction. As Maher argues, we may expect that a good method of theory construction produces theories which are predictively successful with a higher probability than theories produced by a bad method. The hypothesis Y^M that the method M that has been deployed in the construction of a theory H was good thus predicts that the rate of novel predictive success of theories produced by M will be above average. In other words, it predicts that among the novel predictions of theory H, that is among those predictions of data E by theory H for which $O^H(E)$ is true, there will be a particularly high ratio of instances where $(E \wedge O^H(E))$ is true. Any instance of $(E \wedge O^H(E)$ then will in turn increase the probability for Y^M being true. This in turn implies that instances of novel confirmation of theory H increase the probability of the truth of so far untested predictions of theory H, which increases the probability of the truth of H. Therefore, $O_H(E)$ contributes to the confirmation of H.

Let us now look at the case of accommodation, where $\not{O}^H(E)$ is true. No scientifically coherent method, be it good or bad, can generate a theory based on data E that contradicts E. Maher does not address the question as to whether the observation that a theory was generated at all tells something about the quality of the method applied (One might imagine that bad methods fail more often to generate any theory that agrees with the data than good methods). Instead, he focuses on the point that instances of accommodation to data E do not constitute examples of method M generating successful predictions. Therefore, he argues, instances of accommodation do not increase the probability of the truth of Y^M and

no higher truth probability for H can be inferred. Maher's argument therefore demonstrates a confirmation extra value of novel confirmation over accommodation.

Other authors make alternative suggestions regarding the nature of the statement Y that predicts novel confirmation. Kahn et al. (1992) propose that the occurrence of novel confirmation is taken by the scientific community as an indication of the high quality of those scientists who developed the corresponding theory. Scientists of high quality are taken to have a tendency of developing predictively successful theories. Each instance of novel predictive success thus raises the trust in the capabilities of the scientists who developed the corresponding theory. A high quality of the involved scientists then in a second step is linked to a higher probability of the truth of theories they construct. This finally raises the trust in the truth of a predictively successful theory beyond the level that had been achieved based on the theory's agreement with the empirical data.

Barnes (2008) agrees with Kahn et al. on the point that novel confirmation is relevant primarily for assessing the quality of involved scientists. He differs from their account in two important ways, however. First, he argues that the crucial criterion for novel confirmation must be theory endorsement rather than theory building. Scientists may come up with a theory without actually believing it. Barnes argues that it would be implausible to relate these scientists' qualities to the appraisal of the theory they developed but misjudged. According to Barnes, novel confirmation thus should be taken to apply if a theory has been *endorsed* by a scientist without relying on the confirming data. Second, Barnes emphasizes that a scientist's endorsement of theories which then turn out to be empirically successful should be understood to be based on her true background beliefs: scientists endorse a theory based on their background beliefs. If those beliefs are largely true, they will tend to endorse theories which later turn out empirically successful. Novel confirmation thus is an indicator of true background beliefs of those scientists which endorsed the theory. The scientists' abilities are thereby connected back to objective facts about the world.

A quite different perspective on novel confirmation is developed by Hitchcock and Sober (2004). They argue that some, if only limited, advantage of novel confirmation over accommodation can be extracted from an analysis of curve fitting. If one tries to fit given data-points by a freely chosen function, there is a risk of mistaking purely statistical aspects of the distribution of data points for structural characteristics of the true generating function of those data points. This can lead to a sub-optimal selection of fit-functions if one does not account for that danger by adequate penalty terms for complexity (Forster and Sober 1994). It can be shown that a function that avoids overfitting provides more accurate predictions of future data points than a function that involves overfitting. On that basis, novel confirmation can be used as an indicator that the chosen function actually reproduces structural characteristics of the true generating function and over-fitting has been avoided. Novel confirmation thus gives extra information about the quality of a fit function that goes beyond what can be provided by accommodation.

Though Hitchcock and Sober differ significantly from the approaches presented before, their rationale adheres to the same general scheme of reasoning. In their case, the meta-statement Y^{HS} is the claim that over-fitting has been avoided when constructing H. Y^{HS} predicts that higher predictive success rates will be achieved. Any instance of predictive success confirms Y^{HS} and thus also raises the probability of future predictive success of theory H.

3: Limitations to Scientific Underdetermination and Novel Confirmation

Let us once more repeat the general structure of the novel confirmation arguments reconstructed in Section 2. We select a theory H and a hypothesis Y about theory H so that the following conditions hold:

1) The truth (or empirical viability within a certain regime) of H cannot be univocally determined based on the available data.
2) The truth of Y cannot be univocally determined based on the available data.
3) If Y is true, H has a higher probability of being true (or empirically viable within a certain regime) than if Y is false.
4) The observation $O^H(E)$ that novel confirmation of H by empirical data E has occurred raises the probability that Y is true.
5) The observation $\not{O}^H(E)$ that accommodation of empirical data E to H has occurred does not raise the probability that Y is true (or raises it to a lesser degree than $O^H(E)$).

From 1),..,5) it follows that novel confirmation raises the probability of H being true to a higher degree than accommodation. We thus have a confirmation extra value of novel confirmation over accommodation.

I now propose the following new candidate for statement Y:

> $Y=L_n(E_0,S_1)$: There are no more than *n* different possible scientific theories which are in agreement with the available data E_0 and can be distinguished by some not yet realized set of experiments S_1.

To make this statement meaningful, two kinds of clarification are necessary. We must define what counts as a possible scientific theory and we must introduce rules for the individuation of theories within that framework. I will postpone those clarifications to the next section, however, and first present the basic structure of the argument. To that end, I will first motivate the proposed choice of hypothesis Y and then show that, provided that it can be turned into a meaningful statement, it satisfies conditions 2) to 5) for theories H that satisfy condition 1).

Let us begin with some general thoughts about the role of alternative scientific theories in the scientific process. Usually, we take scientific theory building to be underdetermined by the available empirical data: we tend to think that we could, in principle, construct a number of scientific theories which are compatible with the available data. We may have found some of those theories, or maybe just one of them, but we assume that as yet 'unconceived alternatives' (Sklar 1980, Stanford 2006) exist as well. If we assumed that the underdetermination of scientific theory building were entirely unlimited, however, we could not explain why novel predictive success arises in science at all. To understand this important point, let us imagine the following scenario: we have collected empirical data E_0 in a set of experiments S_0 and have developed a theory H that can account for that data. Theory H makes further predictions which we intend to test in a new set of experiments S_1. Now let us assume that theory building is entirely unlimited, that is, any imaginable outcome of the upcoming experiments S_1 can be accounted for by a satisfactory scientific theory. If that were the case, why should we believe that the predictions with

respect to the outcomes of S_1 provided by the theory H we have actually developed will turn out true? Given that we have developed theory H just on the basis of data E_0 and do not have a truth detector that guides us towards the true theory, what reasons do we have for believing that ours rather than one of the uncounted empirically different alternative scientific theories makes the correct predictions?

Actually, scientists often do find themselves in that kind of situation and indeed see no reason for trusting their theory. (In particular they may have developed many theories already and have no reason for trusting one rather than the other.) There are other contexts, however, where scientists do trust the predictions of their theories. In order to trust those predictions, scientists must believe that the number of possible alternative theories is, in some sense, significantly limited. If there are only few scientific theories which can account for data E_0, then chances are that the one scientists have developed will turn out to be true, or at any rate empirically successful at the next stage of experimental testing. If scientists can infer limitations to scientific underdetermination from predictive success, however, this means that a hypothesis on limitations of scientific underdetermination in the given context satisfies condition 3) of the five conditions stated above. Such a hypothesis therefore looks like a natural candidate for Y. $L_n(E_0,S_1)$ constitutes a straightforward form of a claim of limitations to scientific underdetermination. So let us try to develop a specific model on its basis.

Before entering the details of the analysis, it is important to make one remark on the definition of confirmation in our context. In Bayesian terms, confirmation of theory H is usually taken to correspond to a rise in the probability of the truth of H. Since the absolute truth of theories raises problems in the context of physics – for example, physicists often confirm theories which are known to be strictly speaking false since they are inconsistent beyond a given regime – it seems advisable to relate confirmation to a more modest concept than truth, that is to empirical viability within a given regime. I call a theory empirically viable within a given regime if it can reproduce all empirical data within an intended regime of empirical testing. Non-relativistic quantum mechanics is empirically viable in this sense if it can reproduce all empirical data in microphysics in the regime where relativistic effects can be neglected.

We want to consider a series of generations of experimental testing S_i , i=0,1,2.., where each new generation of experimental testing S_{i+1} includes the earlier generation of experiments S_i. The observation that E_i has not entered the construction process of H is called $O^H(E_i)$. $n(E_i,S_j)$ denotes the number of possible alternative theories which are consistent with empirical data E_i and can be distinguished by experiments S_j, where i<j.

We now introduce a theory H that has been developed based on empirical evidence E_0 that was collected in a set of experiments S_0. Theory H accounts for E_0 and makes non-trivial predictions regarding the outcome of a future set of experiments S_1. Scientists then carry out experiments S_1 and find that the collected evidence E_1 is consistent with the predictions of H.

In order to discuss this process, we want to assume the following model. There are a finite number $n(E_0,S_1)$ of possible theories which can be constructed in agreement with E_0 and can be empirically distinguished by experiments S_1. Each theory that is compatible with E_0 has the same subjective probability[1] of being viable, irrespectively of the question whether or not it has actually been developed by scientists. In other words, scientists have

[1] Subjective probabilities denote the probabilities we attribute to the theories based on our understanding of the overall situation.

no other way than agreement with empirical data to assess the quality of a theory. Inversely, we can then conclude that, if *n* possible theories exist and scientists find one theory H that accounts for the available data E_0, the chances of that theory being viable with regards to a set of characteristic experiments S_1 is

$$P\left(T_1^H \middle| O^H\left(E_1\right), \wedge \Theta^H\left(E_0\right)\right) \cong 1/n(E_0, S_1) \quad (2)$$

where T_H^1 denotes the claim the theory H is empirically viable with regards to its characteristic experiments S_1.

We thus have established a significant connection between the number of possible alternative theories *n* and the probability of novel predictive success of a given theory. We can then make the following statement: $L_n(E_0,S_1)$ with a low number *n* can explain the novel predictive success of theory H since, under the condition that $L_n(E_0,S_1)$ is true, there is a significant probability that theory H is empirically successful. As long as no other equally satisfactory types of explanation of the novel predictive success of H are available (that is, no explanations which do not rely on limitations to scientific underdetermination), inference to the best explanation can lead to the conclusion that some $L_n(E_0,S_1)$ with low *n* is probably true. In other words, we can infer from novel predictive success of theory H that underdetermination is limited. Novel confirmation with respect to data E_1 raises the probabilities of the truth of statements $L_n(E_0,S_1)$ with low *n*.

Now let us consider another set of characteristic novel predictions of theory H that can be tested by a new set of experiments S_2. Just like the predictions with respect to S_1, these predictions are provided by theory H which resulted from the scientist's search for a coherent theory that could account for data E_0. Given that we deal with core implications of the same theory and remain within the same research context, we have no reasons to assume (in the absence of additional information) that the spectrum of possible alternative theories to H which give different empirical predictions with respect to S_2 will be radically wider than the spectrum of alternative theories that give different predictions with respect to S_1. The limitations to scientific underdetermination established with respect to theory individuation based on S_1 therefore should be roughly applicable with respect to E_2. We are led to assume that high probabilities for statements $L_n(E_0,S_1)$ with low *n* imply similarly high probabilities for the corresponding low *n* statements $L_n(E_0,S_2)$ and therefore also high probabilities for low *n* statements $L_n(E_1,S_2)$.[2] A theory that has provided successful predictions with respect to data E_1 thus is more likely to provide successful predictions with respect to empirical data E_2 in the future than a theory that has not provided successful predictions and therefore provides no basis for believing in $L_n(E_1,S_2)$ with low *n*. Since empirical data that is predicted by theory H confirms H, this means that novel confirmation raises the probability that theory H will be empirically confirmed in experiments S_2. Thereby, it raises the probability that H is true or empirically viable within an intended regime.

Now let us compare this situation with the case of accommodation of data E_1, expressed by $\Theta^H(E_1)$. $\Theta^H(E_1)$ implies that H agrees with data E_1. We thus have:

$$P\left(T_1^H \middle| \Theta^H\left(E_1\right)\right) = 1$$

[2] Obviously, we have $n(E_0,S_2) > n(E_0,S_1)$ since S_2 includes S_1; and $n(E_0,S_2) > n(E_1,S_2)$ since E_1 includes E_0.

The fact that H is in agreement with data E_1 thus is fully explained by $\not{O}^H(E_1)$. Statements of the kind $L_n(E_0,S_1)$ thus cannot contribute to the explanation of $\not{O}^H(E_1)$. Thus, there is no basis to infer a small number of possible theories $L_n(E_0,S_1)$ from the fact that H agrees with E_1. This implies that there is no basis either to draw any conclusions with regard to $L_n(E_1,S_2)$. Accommodation does not provide any confirmation extra value via the assessment of the number of possible scientific theories. It follows that novel confirmation by data E_1 provides stronger confirmation of theory H than accommodation to data E_1.

4: Specifying the Number of Alternative Theories

So far, we have used the statement $L_n(E,S)$ without offering a solid basis for its meaning. It is clear that, without a rigid framework that determines what counts as a scientific theory in a given context, a statement on the number of possible scientific theories must remain meaningless. The required framework can be established by defining constraints C that have to be satisfied by scientific theories. In order to be effective, the constraints C must allow scientists to be very confident that there is a scientifically viable solution to the problem they are working on that satisfies C. If one could not be very confident that *some* theory that satisfies C is scientifically viable in the intended regime, the number of theories which do satisfy C could not tell much about the empirical viability of an individual theory H from this group. Thus, the question arises: what kind of condition justifies such confidence?

Conditions which are related to a specific scientific strategy or to specific conceptual presumptions always run the risk of being toppled at the next step of scientific progress and therefore don't seem advisable. Rather, C should be constituted by very basic 'scientificality' conditions. There should be a scientific consensus in the field that a theory that violates any of the conditions C would not be accepted as a scientific theory. The set of scientificality conditions may be expected to include requirements of consistency, the absence of ad-hoc assumptions to explain individual events out of a large ensemble, and a certain degree of universality. In contemporary fundamental physics, these three conditions arguably provide a workable framework for theory development. It seems fair to say that any formal solution to a problem that fulfils the three conditions (or just gives reasons to believe that it can fulfil them) counts as a serious scientific contender. In less formal fields where the concept of consistency (sometimes even the notion of consistency with the data) is more difficult to define, other field-specific conditions have to be added.

It is important to note that, even though a specific number *n* only makes sense with respect to specific conditions C, a precise specification of the scientificality conditions is not necessary for giving explanatory extra value to novel confirmation based on our argument. All that is necessary is to be very confident that the true theory fulfils *some* well-defined set of conditions C which are also adhered to by the scientists who develop theories in the field. Estimating the number of possible alternatives then means estimating the number of all alternative theories which satisfy these conditions C. Since our estimate of that number relies entirely on the observation of novel confirmation rather than on an analysis of the constraints C themselves, a fuzzy understanding of C does not impede the analysis.

It remains to be clarified how to individuate theories with the framework of scientificality conditions C. Before entering the details of this discussion, one important point should be emphasised. It is the goal of this analysis to determine the kind of information which can be extracted specifically from novel confirmation and which can eventually be used for raising the probability of the empirical viability of the given theory. We aim at

supporting the following claim: given the chosen principles of theory individuation, novel confirmation can provide information about the number of possible theories. A claim of this kind will always be embedded within a specific scientific context. We do not need to look for a principle of theory individuation that is universally adhered to in science. Since intuitions on what counts as an individual theory vary considerably in science, such a universally valid principle would be impossible to find. It is important, however, that the presented principles of theory individuation are not alien to science and are applied in relevant scientific contexts. Examples of specific scientific contexts will be given along the way in order to demonstrate that this condition is satisfied.

So what can we say about the specifics of theory individuation? Since we are interested in predictive success, we want to distinguish theories by their empirical implications. Therefore, we count theories separately only if they give different empirical predictions.[3] Not every theory that is in principle empirically distinguishable from H is of relevance for understanding an individual instance of novel confirmation, however. In Section 3, we have discussed novel confirmation within the context of solving a specific set of problems related to the task of fitting some empirical data E_0. Once scientists have found one solution H to that set of problems, the prospects of the empirical success of theory H with respect to its core predictions E_1 depended on the number of alternative solutions to the specific set of problems related to fitting data E_0. Conceptual alternatives which were based on the same solution to these problems and could only become empirically distinguishable beyond the range of its characteristic predictions E_1 did not affect the chances of predictive success within that context.[4] Therefore, we only counted theories as alternatives which constituted autonomous solutions to the set of scientific problems that were to be solved in the given context. Theories which could differ only with respect to their more far-reaching implications but shared the same effective theory that could already be understood as a solution to the given set of problems did not count as different theories.

Having thus provided the basis for distinguishing empirically distinct theories from theories which are empirically undistinguishable within the given empirical regime, we still have to draw the distinction between empirically distinguishable theories and empirically distinguishable specifications of one theory. The adequate way of drawing that distinction depends on the level on which the phenomenon of novel prediction occurs in a given context. Let us illustrate this by looking at the example of the prediction of the Higgs particle by the standard model of particle physics before the discovery of the Higgs particle in summer 2012. The standard model at the time predicted the existence of a Higgs particle with a mass within a specific range of values. This was the level at which the standard model prediction was taken seriously. Therefore, it makes much sense to count as one theory the general scheme of the standard model that provided the described prediction. Let us compare this choice of theory individuation with a different one where we would insert precise values for the Higgs mass and count each realization of the standard model with a specific value of the Higgs mass as an individual 'theory'. This strategy is inadequate because no physicist would have taken the specific predictions of these individual 'theories' seriously.

[3] While some philosophers and participants to the foundational debate of quantum mechanics don't adhere to this principle of theory individuation, it seems fair to say that it is endorsed in most parts of empirical science: theories are taken to be genuinely distinct only if they can be distinguished by empirical means.

[4] To give an example from actual science, the chances of success for the predictions of the standard model of particle physics, which aims at describing nuclear interactions, are not affected by the range of possible solutions to the problem of the unification of nuclear interactions and gravity. Theories concerned with the latter problem have their characteristic predictions at a much higher energy scale and therefore remain irrelevant for the phenomenology predicted by the standard model.

Before the Higgs mass was measured experimentally, there was no reason to believe in one of those 'theories' rather than another. Therefore, the experimental measurement of a Higgs mass of 125 GeV was not understood in terms of a successful novel prediction of the corresponding 'theory' that 'predicted' that mass value. Rather, it was understood as an empirical specification of a parameter value that had not been predicted by the theory itself. Counting each value for the Higgs mass as an individual theory thus makes no sense because it does not reflect the actual character of the scientific prediction in the given case.

Based on the presented principles of theory individuation, we can now understand $L_n(E_0,S_1)$ as the statement that there are no more than *n* theories which satisfy scientificality conditions C, can account for data E_0 and can be empirically distinguished based on experiments S_1.

The question remains as to whether the model presented in Section 3 is sufficiently realistic as a model of scientific praxis. One may object to this model by arguing that actual scientific reasoning can be influenced by additional principles and therefore does not accord with the assumption that scientists can infer a theory's quality only from the theory's agreement with empirical data. In this vein, it has been suggested (see e.g. Boyd 1990) that, among a number of consistent scientific solutions, the simpler or more beautiful solutions have significantly higher probabilities of being true or viable.[5] In its general form, this objection does not threaten our approach, however. If we assume that scientists are guided by principles of simplicity and beauty, it may be expected that the relevance of those principles increases the probability that scientists find viable theories. As long as the effect is not too strong, it suffices from our perspective to state that $P(T_1|O_H(E)) > 1/n$ rather than $P(T_1|O_H(E)) = 1/n$ must hold under the given circumstances and the claim on limitations to scientific underdetermination remains intact. If one is very convinced that a certain level of simplicity or beauty *must* be met by the empirically viable theory, the corresponding conditions can be included in the conditions C and the entire analysis can be carried out on that basis. In that case, criteria of simplicity or beauty do not replace the assessment of limitations to scientific underdetermination but merely modify the framework of that assessment.

The only way to deny the relevance of assessments of limitations to scientific underdetermination in the given context would be to claim that scientists can be sure always to find the simplest and most beautiful theory first and that the simplest and most beautiful theory is always true. This, however, seems clearly contradicted by the historical record in science. A low number of alternative theories therefore can provide a fairly satisfactory explanation of novel predictive success of a corresponding theory H under very plausible assumptions even if elegance or simplicity do increase a theory's probability of being true.

5: Comparison with Other Approaches

One crucial point distinguishes the approach of limitations to underdetermination from the conceptions of Maher, Kahn et al. and Barnes. The latter assume that novel confirmation tells us something about the quality of the contingent processes of theory generation or theory endorsement. According to them, we can learn from novel confirmation about the

[5] I have serious doubts about the deployment of principles of simplicity and beauty in the given context and do not endorse that approach myself. The paragraph just aims at demonstrating that, if the approach makes sense, it is compatible with the approach of limitations to underdetermination.

quality of the scientists who happened to be concerned with the theory or about the quality of the method that happened to be followed when the theory was created. The approach of limitations to underdetermination, to the contrary, is based on the idea that novel confirmation tells us something about the space of possible alternative theories. Novel confirmation thus is taken to clarify the 'objective' scientific framework within which the theory is developed.

Despite that substantial difference, the approach we propose is conceptually related to Maher's. We want to clarify this relation by modifying Maher's approach in a way that leads towards the question of limitations to underdetermination. Two steps of modification are necessary to that end. First, we have to place Maher's method M at a more fundamental level than intended by him and take it to be the scientific method itself. On that basis, of course, Maher's idea to distinguish individual instances of theory generation based on the specific methods that were applied does not work anymore. After all, any serious scientist uses the 'scientific method'. In order to make use of Maher's approach nevertheless, we therefore have to compare scientific problems rather than theories which aim at solving the same problem. We then assume that the scientific method is more likely to lead to true (or empirically viable) theories in some contexts than in others. According to our modified Maher-approach, novel confirmation tells us that we are facing a scientific context where the scientific method works well in the stated sense. But why is it that the scientific method is more likely to provide true or viable results in some specific contexts than in others? In order to answer this question, we introduce our concept of limitations to scientific underdetermination: in some contexts, the scientific method provides a framework where only few theoretical solutions are possible, which implies that the solution the scientists come up with has decent chances of being viable.

Method thus plays an important role in Maher's approach as well as in the approach of limitations to underdetermination. In the latter, however, it merely provides a framework of reasoning and does not do the work of distinguishing between different theories. This reduced role of method in the approach of limitations to underdetermination avoids criticism that has been levied by several authors against Maher's deployment of method. These lines of criticism (see e.g. Howson & Franklin 1991 and Lange 2001) have emphasised the ambivalence of the concept of method deployed by Maher: it seems unclear whether it is possible to identify methods within the scientific process which show the formal characteristics attributed to method in Maher's model; and it seems difficult to formulate a plausible and generally applicable conceptual distinction between method and theory in scientific contexts.

Marc Lange argues that the apparent intuitive strength of Maher's presentation of the method concept is due to his choice of a specific example that does not represent a genuinely scientific scenario at all. Maher presents as his core example a scenario where two 'scientists' are asked to write down the results (heads or tails) of a series of 100 coin tosses which are to be carried out in an experiment. The first 'scientist' S1 sees 99 tosses and writes down the entire series afterwards. The second 'scientist' S2 sees only 50 tosses and then writes down the entire series. Maher considers the case where S2 gets her predictions of tosses 51-99 exactly right. He plausibly argues that, after having seen 99 tosses and the statements of S1 and S2, we would give the prediction of S1 regarding toss number 100 just a probability of 50% while we would give the prediction of S2 a probability that is much higher. The reason for that, Maher claims, is that we infer from the novel confirmations in the case of S2 that S2 has a good method of generating her predictions. In the case of S1 who just accommodated data known to her, we have no reason for making that inference.

Novel confirmation thus may be understood to confirm a theory (in the given case S2's statement of the full sequence) based on an inference to the quality of the method that generated the theory.

Lange criticises that this scenario does not exemplify a comparison between two alternative methods of generating a scientific theory. Rather, novel confirmation is required in this scenario for establishing that the 'scientist' has a reasonable method at all and is not just guessing. In the absence of novel confirmation, one would assume that the coin tossing is a random process and no viable theory exists that predicts its outcome. We would thus assume that any prediction amounts to mere guesswork. Only strong examples of novel confirmation can alter our position in specific cases. To the contrary, actual scientific reasoning can be expected to aim seriously at understanding the law-guidedness of the process under investigation. The question as to whether or not we should attribute a serious scientific method to the scientist who develops a certain theory thus does not arise. Maher's account fails on a general basis, according to Lange, because the distinction between method and theory is only clear as long as 'theory' is understood in the simplistic way exemplified in the coin toss example. There, the theory is a simple data model and all that is conceptual is attributed to the method. We can only see the data model and the conceptual basis (the method) remains entirely hidden to us. In genuine examples of scientific reasoning, however, theories are highly complex and multi-layered concepts themselves. Moreover, the conceptual basis is openly explained to the scientific community, which can judge the quality of both theory and method based on that information. It thus seems far less obvious whether and how theories can be clearly distinguished from methods and whether much can be gained from the distinction between method and theory with respect to understanding the role of novel confirmation.

The approach presented in this paper can be understood as a way of avoiding the described problem. The method of theory building is identified with the scientific method itself. The scientificality conditions that define the scientific method are so general that they can be clearly distinguished from the principles which define a specific theory within its framework. Moreover, the 'scientific method' is shared by all scientists in the field and thus does not bear the responsibility of distinguishing between more and less promising theories. That work is done by the limitations to scientific underdetermination, which, as I hope to have demonstrated in the previous section, can be defined with more clarity than the quality of a method.

Greater conceptual clarity may be seen as a general advantage of the approach of limitations to scientific underdetermination over those of Kahn et al. and Barnes as well. The scientist's capability as used in Kahn et al. is a rather general and vague concept whose explanatory power may be taken to suffer from that vagueness. Barnes tries to fill that concept with meaning by relating it to correct background assumptions, which, however, remain a fairly vague concept themselves. The notion of possible alternative scientific theories arguably constitutes a more specific and therefore more powerful explanatory tool.

6: The Case of Fundamental Physics

For a number of reasons, the approach of limitations to underdetermination appears more convincing than the approaches of Kahn et al., Barnes and Hitchcock and Sober in the context of fundamental physics and comparable disciplines. In order to demonstrate this, we

first have to analyse the wide range of problems faced by the cited approaches in the specific context of fundamental physics.

Let us start with another look at Hitchcock and Sober's idea to reduce the significance of novel confirmation to testing the degree of over-fitting. Hitchcock and Sober concede in their paper that their analysis does not necessarily allow for a straightforward extrapolation towards more complex scientific theory building. In the context of fundamental physics, Hitchcock and Sober's approach indeed seems largely inadequate. While scientific analysis based on curve-fitting does occur in some contexts – good examples are attempts to find long term regularities of sun-spot activity or large scale structures of the universe – much of fundamental physics happens at the other end of the spectrum: fundamental physical theories tend to be developed within complex mathematical frameworks where consistency requirements strongly constrain the options of coherent theory building. The most impressive examples of predictive success in fundamental physics occur in highly constrained regimes where the significance of novel confirmation cannot be related to the question of over-fitting at all.

This can be shown with particular clarity in the case of the successful prediction of light bending by Einstein's theory of general relativity, which arguably constitutes one of the most famous examples of novel confirmation in science. General relativity was developed in order to reconcile the principles of special relativity with gravitation. During the ten years Einstein worked on solving the conceptual problems which arose on the way towards reaching his theoretical goal, no specific empirical data entered the process of theory development. Thus no-one could reasonably doubt the viability of general relativity based on the claim that it was over-fitting empirical data. The situation in quantum mechanics or gauge field theory is quite similar. Successful novel predictions are largely based on the deployment of general physical principles which offer solutions to consistency problems that have arisen at a different level of analysis than data fitting. Over-fitting of data does not constitute a relevant question in those contexts. It seems fair to say that Hitchcock and Sober do not address those aspects of novel confirmation which are most crucial in fundamental physics.

A similar verdict seems justified with respect to Kahn et al. (1992). The conception of individual scientists who offer their own theories on a subject and have to be evaluated in order to determine their theories' prospects may be applicable in some cases but seems far-fetched as a general characterization of the field. The most influential physical theories like quantum mechanics or the standard model of particle physics – though first suggested in their basic outlines by individual scientists or small research groups – later turned into multi-faceted joint enterprises of concern to the entire research field. Novel confirmation of those theories often happened after that theory had become the dominant scientific research program in the field worked on by large numbers of scientists and research groups of various qualities. More often than not, no alternative theory capable of solving the same theoretical problems was known at the time. In this vein, it does not make much sense to understand the significance of novel confirmation in terms of its role as a means of assessing the quality of individual scientists or research groups. If at all, one would have to understand the assessment as referring to the scientific abilities of the physics community as a whole. It is quite difficult to understand, however, what could be meant by that.

Barnes' approach seems better equipped to deal with that specific criticism. In his understanding, novel confirmation does not determine individual capacities of physicists. It rather determines the quality of the background knowledge that leads to endorsing a certain theory. If many or all scientists in the field share the same background knowledge, they may

plausibly endorse the same theory on that basis. It would still make sense in that case to test the involved background knowledge by looking at novel confirmation.

Barnes' approach looks problematic for a different reason, however. Barnes focuses on theory endorsement rather than theory construction and assumes that the time of theory endorsement is the crucial criterion for distinguishing between old and novel confirmation. This forces him, however, to assume that the first endorser's analysis – which cannot rely on other endorsements - cannot itself rely on arguments of novel confirmation. Barnes offers no convincing reasons for the a priori implausible assumption that arguments related to novel confirmation play no role for the first endorser. In order to see the problem, let us assume a theory that has been developed by some scientists but has not been endorsed by them or anyone else because it makes highly speculative and counter-intuitive claims. Let us further imagine that some of the theory's novel predictions then get confirmed. Most adherents to novel confirmation would take such a scenario as a classical scenario where novel confirmation has extra confirmation value: the fact that the theory gets predictions right suggests that, despite its counter-intuitive claims, it is on the right track. For Barnes, however, the described scenario cannot be an example of novel confirmation because no theory endorsement has happened before confirmation.

In the context of fundamental physics, Barnes' focus on theory endorsement looks clearly inadequate. As an example, let us once more consider the case of the discovery of the Higgs particle. Since the Higgs particle constitutes a core prediction of the standard model, it seems like a perfect example for novel confirmation. Following Barnes' scheme, novel confirmation of the standard model should have arisen with regard to the Higgs data because it raised the trust in those scientists who had endorsed the Higgs hypthesis already before it had been empirically confirmed. This, however, neither makes scientific sense nor agrees with what actually happened. Next to all high energy physicists endorsed the Higgs hypothesis in recent decades. So, the only way to find a distinction in endorsing it would be to go back to the early days of the standard model. It would be absurd, however, to ground the extra value of novel confirmation in the Higgs case on the assessment of those scientists who endorsed the Higgs hypothesis in the 1960s or 1970s. The conceptual understanding of standard model physics is vastly superior today than it used to be 40 years ago. Today's analysis of the standard model, however, is mostly carried out by physicists who had not even started their career in the 1970s and thus have no record of early endorsement of the standard model. Moreover, most physicists today have no specific knowledge about the historic details regarding endorsements of the standard model in the early 1970s. What every particle physicist does know and what does matter for theory assessment is that the theory itself has an impressive record of novel predictive success. Novel confirmation quite obviously is understood to tell something about the standard model and, more specifically, about the Higgs hypothesis itself rather than about the physicists who had endorsed it before its predictions had been confirmed.

I want to point to another problem that arises for Hitchcock and Sober, Kahn et al. and Barnes. The problem, which I will call the problem of inherited merits, is related to a somewhat surprising and seemingly paradoxical aspect of the scientists' intuitions regarding novel confirmation. Though the problem is not confined to fundamental physics, it can be nicely observed in that context. Let us imagine a scientist A who develops and endorses theory T_A that correctly reproduces the data E_0. Let us assume that T_A predicts data E_1, which is later confirmed in an experiment S_1. Let us further assume that, after E_1 has been collected, a scientist B understands that E_0 and E_1 can also be reproduced by an alternative theory T_B, which, however, makes different predictions than T_A regarding a new experiment

S_2. The question now arises: should T_A be taken to have a higher probability of being true (or empirically viable at the next experimental step) than T_B on the grounds that T_A, unlike T_B, has been confirmed by novel data? The scenarios presented by Hitchcock and Sober, Kahn et al. and Barnes 2008 would suggest that it should. Following Hitchcock and Sober, predictive success indicates the absence of over-fitting in T_A while a theory developed after the evidence E_1 has been taken into account runs the risk of over-fitting once again. T_A thus is favored over T_B. According to Kahn et al. as well as Barnes 2008, novel confirmation establishes that A is an able scientist with sound background knowledge who is capable of developing respectively endorsing viable theories. With respect to scientist B, this has not been established. The different assessments of scientists A and B then enter the assessments of the chances for the viability of T_A and T_B and thus favor T_A.

In an actual physical scenario of that kind, the situation would be different, however. Admittedly, it may be the case that third party scientists make their first decision whether or not to study theory T_B based on assessing the merits of its creator or endorser B. Once they have studied the theory and have acknowledged that it is coherent and indeed capable of reproducing E_0 and E_1, however, the identity of physicist B would (more or less) drop out of the equation. Theory T_B would be taken to be just as good a candidate for being viable as T_A.

A good example of a situation of this kind is the idea of large extra dimensions in high energy physics that became popular in recent years. For a long time, descriptions of microphysics were based on the notion that space-time was four-dimensional all the way down to the Planck length. Phenomena like the unification of the three gauge couplings at a high energy scale (the so called GUT scale) were explained by theoretical concepts that were based on the assumption of four-dimensional space-time. In the 1990s it was then understood that the phenomenology could also be explained by introducing large[6] extra-dimensions. In the corresponding models, space-time is taken to be of a dimension higher than four at small distances. Four dimensional space-time only emerges at distances above the electroweak scale (regarding nuclear interactions) respectively beyond the scales tested by precision measurements of gravitation. Above those scales, the additional dimensions are not observed either because the respective dimensions are compactified with a small radius or because they have a very specific geometric structure (constituting so called warped dimensions). The crucial question for physicists assessing the plausibility of the new models is whether these models can reproduce the phenomenology just as well as the more traditional theories. The new models are taken to be equally viable candidates if they stand that test. The predictive successes of the old 4-dimensional models are not understood to give them an advantage over the new hypotheses.

At this point, Hitchcock and Sober, Kahn et al. and Barnes could still try one line of reasoning. They could say: in the given case, theory evaluation apparently does not rely on assessments of over-fitting, the quality of involved scientists or their background knowledge. Therefore, novel confirmation does not seem to be significant. That understanding, however, does not accord with their own account of the role of novel confirmation. In the case of Hitchcock and Sober, any curve fitting scenario allows for later accommodation of a new curve to an extended data set. If novel confirmation did not play any role in such scenarios, it could not play a role in curve fitting at all. In the cases of Kahn et al. and Barnes, it seems obvious that the information gained due to novel confirmation about the scientist who developed or endorsed the corresponding theory cannot be devaluated by the mere

[6] ‚Large‘ still means quite small in this context. Large extra dimensions are taken to be significantly larger than the Planck length but too small to be observable in experiments up to this point.

fact that someone else developed a different theory based on accommodation later on. Why should our opinion about one scientist's quality be altered by the information that different scientist developed a different theory? A satisfactory philosophical analysis of the novel confirmation problem thus cannot deny the relevance of novel confirmation in cases where the phenomenon of inherited merits applies but rather must offer an explanation of how the mechanism of inheritance works in the presence of novel confirmation.

Explaining novel confirmation in terms of assessments of limitations to scientific underdetermination avoids the problems described above.

Let us first look at the inadequacy of understanding novel confirmation in fundamental physics in terms of curve fitting. The approach of limitations to underdetermination is tailored to account for the scenarios we find in fundamental physics. Rather than assessing the degree of over-fitting of a scientific theory to empirical data, it assesses the degree to which the conditions of scientific reasoning exclude data patterns which might seem possible alternatives at first sight. The process of extracting theoretical conclusions from consistency arguments in this understanding gains credibility from the understanding that scientific underdetermination is substantially limited. This understanding, in turn, is inferred from novel confirmation.

Attempts to relate novel confirmation to assessments of the qualities of scientists have been shown to be implausible due to the long time scales between theory creation and theory confirmation in high energy physics as well as due to the fact that crucial theories often cannot be attributed to individual scientists or research groups. Relating the question of novel confirmation to limitations to scientific underdetermination reflects this situation by supporting the understanding that novel confirmation tells something about the theory's conceptual embedding rather than about scientists. Furthermore, the approach of limitations to underdetermination focuses on theory creation rather than theory endorsement. The problems related to Barnes' focus on theory endorsement therefore are avoided as well.

The extent to which limitations to scientific underdetermination can account for the way novel confirmation is treated in fundamental physics can best be appreciated by looking at the problem of inherited merits. Limitations to scientific underdetermination provide a clear solution to this problem. The predictive success of theory T_A with respect to data E_1 is taken to indicate limitations to the number n of scientific theories which can be constructed based on the data E_0. These limitations eventually indicate the chances of T_A for offering correct predictions of data E_2. Since T_B and T_A are both elements of the very same set of scientific theories which are compatible with E_0 and E_1, all statements on limitations to scientific underdetermination relevant for T_A must be equally relevant for T_B as well. Therefore, the 'inheritance of merits' is strictly implied by the approach of limitations to scientific underdetermination.

Another tricky aspect of the 'inherited merits' scenario looks entirely straightforward from the perspective of limitations to underdetermination. It is natural to acknowledge that the discovery of T_B has a certain detrimental effect on the expectations regarding the predictive success of T_A. After all, its discovery opens up the explicit possibility that T_B rather than T_A will be predictively successful with respect to experiment S_2. While considerations of that kind are not incompatible with other attempts to explain novel confirmation, they do not fit in very well either. They constitute an additional argument quite alien to the line of reasoning deployed when establishing the significance of novel confirmation. Once novel confirmation is related to limitations to scientific underdetermination, however, all

considerations neatly blend into one coherent whole. The discovery of T_B establishes that a certain degree of scientific underdetermination is realized, while instances of novel confirmation in the given context indicate that scientific underdetermination may be expected to be limited. In conjunction, those items of information create one overall assessment of scientific underdetermination that is in the end responsible for the trust one has in the predictions of T_A and T_B.

7: Conclusion

We have seen that the approach of limitations to underdetermination offers a viable basis for explaining novel confirmation in fundamental physics. The approach clearly does not provide a universal explanation of novel confirmation in all scientific contexts, however. We do find instances of curve fitting in many fields of science where the arguments presented by Hitchcock and Sober apply nicely while the underdetermination approach does not work. Equally, we find contexts in science and beyond where assessments of the capacities of individual scientists or agents play a crucial role in judging a statement's viability. When Warren Buffett tells us about his expectations regarding the stock markets, it seems to make much sense to judge his capability based on his record of novel confirmation and to assess the viability of his statement accordingly. It seems a far less promising idea to evaluate Buffett's statement based on an assessment of the number of possible theories about the stock market.

Still, we have argued in the previous section that the contexts where successful novel predictions occur most conspicuously in science are those where the approach of assessing limitations to scientific underdetermination does apply and other approaches seem less adequate. One might interpret this observation a little more extensively. If we want to understand the particular strength of science in providing successful novel predictions, limitations to scientific underdetermination seem the most promising place to look. The instantiation of such limitations is an essential precondition for the novel predictive success of science and understanding those limitations in a given context is essential for understanding a theory's status and its prospects. On that basis, the strength of limitations to scientific underdetermination constitutes the conceptually most significant thing scientists can learn from novel confirmation. The analysis of limitations to underdetermination therefore can provide the most substantial extra value of novel confirmation.

Acknowledgements: This work was supported by the Austrian Research Fund (FWF) grant P22811-G17.

References:

Barnes, E. C. (2008). *The Paradox of Predictivism*, Cambridge University Press.

Boyd, R. (1990). Realism, Approximate Truth and Philosophical Method. in C. Wade Savage, *Scientific Theories*, Minnesota Studies in the Philosophy of Science, vol 14, University of Minnesota Press.

Dawid, R. (2006). Underdetermination and Theory Succession from the Perspective of String Theory. Philosophy of Science, 73/3, 298-322.

Dawid, R. (2010). High Energy Physics and the Marginalization of the Phenomena. Manuscrito 33/1, special issue *Issues in the Philosophy of Physics*: 165-206.

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

Dawid, R., Hartmann, S. and Sprenger, J. (forthcoming). The No Alternatives Argument, The British Journal for the Philosophy of Science.

Forster, M. R. and Sober, E. (1994). How to Tell When Simpler, More Unified, or Less Ad Hoc Theories Will Provide More Accurate Predictions. British Journal for the Philosophy of Science 45, 1-35.

Hitchcock, C. and Sober, E. (2004). Prediction Versus Accommodation and the Risk of Overfitting. British Journal for the Philosophy of Science 55, 1-34.

Howson, C. and Franklin, A. (1991). Maher, Mendeleev and Bayesianism. Philosophy of Science 58, 574-585.

Kahn, J. A., Landsberg, S. E. and Stockman, A. C. (1992) On Novel Confirmation. British Journal for the Philosophy of Science 43, 503-516.

Lange, M. (2001). The Apparent Superiority of Prediction to Accommodation as a Side Effect: A Reply to Maher. British Journal for the Philosophy of Science 52, 575-588.

Maher, P (1988). Prediction, Accommodation and the Logic of Discovery. PSA 1988, 273-285.

Maher, P. (1993). Howson and Franklin on Prediction. Philosophy of Science 60, 329-340.

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

Sklar, L. (1980). Do Unborn Hypotheses Have Rights? Pacific Philosophical Quarterly 62, 17-29.

Stanford, P. K. (2006). *Exceeding our Grasp – Science, History, and the Problem of Unconceived Alternatives*, Oxford University Press.

Worrall, J. (1978). The Ways in Which the Methodology of Scientific Research Programme Improves on Popper's Methodology. G. Radnitsky and G. Anderson (eds.), *Progress and Rationality in Science*, Dordrecht: Reidel.

Worrall, J. (2002). New Evidence for Old. In P. Gärdenfors, J. Wolenski and K. Kijania-Placek, (eds.), *In the Scope of Logic, Methodology and Philosophy of Science,* 191-209. Dordrecht: Springer.

Zahar, E. (1973). Why did Einsteins's Programme Supersede Lorentz's? British Journal for the Philosophy of Science, 24, 95-123.