THEORY ASSESSMENT AND FINAL THEORY CLAIM IN STRING THEORY

Richard Dawid

String theory has by now maintained a highly influential position in high energy physics for more than a quarter of a century. The present article analyses the reasons for the considerable trust exponents of string theory have in their theory even though it has neither found empirical confirmation nor a complete formulation up to this point. It is argued that this trust can be understood in terms of an emerging new conception of theory assessment that relies strongly on the identification of limitations to the underdetermination of scientific theory building. The second part of the paper makes the point that another conspicuous element of string theoretical reasoning, the prominent role of the notion of a final theory, can be understood in terms of this altered notion of theory assessment as well.

1: Introduction

When string theory was first proposed as a universal theory of all interactions in 1974, it was considered a highly speculative idea with very unclear chances of success. Today, three and a half decades later, the theory constitutes one of the most influential and most intensely analysed concepts in theoretical physics. String theory's ascend has come along, however, with considerable changes of the characteristics of theory dynamics and theory assessment in fundamental physics. Those changes have, in recent years, provoked strong criticism of the string physical enterprise by physicists working in other research fields. In the present paper, we will discuss two controversial aspects of string physics and analyse their philosophical implications.

The first point of controversy is directly related to string theory's peculiar success story since the mid 1980s. Since then, the theory has attracted a large portion of theoretical high

energy physicists and exerts a strong influence on adjacent fields like particle physics model building and cosmology. The theory is widely treated as a well-established and well-trusted theory. While the degree of optimism regarding the theory's prospects has had its ups and downs (after a very optimistic phase in the second half of the 1990s, the last decade has been characterised by a slightly more sober perspective), most string theorists would claim that string theory is likely to constitute an important step towards a deeper understanding of the external world.

Many physicists in other fields have for a long time been irritated by the fact that string physics has acquired its remarkably strong position in high energy physics without having found empirical confirmation or even a complete theoretical formulation. The longer this situation prevailed, the more insisting the voices have become which call this a problematic development that threatens the integrity of the research process in fundamental physics. (for criticism in this direction, see e.g. [Smolin 2006], [Woit 2006] and [Penrose 2005].)

A second worry for many observers is concerned with the final theory claim often expressed in the context of string physics. String theorists at times express their belief that a coherent formulation of string theory – which, to be sure, in many respects may be expected to look quite different from the present incomplete understanding of string physics - would constitute a final theory: no other more fundamental or more universal theory than string theory (or whatever the name given to the theory in then end) would ever be required to account for new empirical data. (Positions along that line have e.g. been expressed in [Weinberg 1994], [Greene 1999], or [Kaku 1997].) Critics consider such final theory claims misguided or even unscientific. One main strand of criticism which will be most relevant for our discussion asserts that final theory claims systematically lack a valid epistemic foundation.

In the following, it will be argued that the two described controversial aspects of string physics are directly related to the same substantial shift in scientific reasoning that is happening in contemporary fundamental physics at this point: the increasing importance of assessments of scientific underdetermination. The disputes between critics and exponents of string physics in this light appear as disputes between defenders of the traditional paradigm of theory assessment and adherents of a newly emerging one. Understanding the character of that new paradigm will be the main purpose of this article.[1] The article does not aim at vindicating the paradigm shift it purports. Any such vindication must in the end be based on the success of the scientific research program that is based on the new paradigm. Two more modest points will be made, however: the new paradigm will be shown to retain the characteristics of scientific reasoning necessary for critical theory assessment – and it will be argued that there are some good reasons for adopting and testing it at the present stage.

We begin our discussion by characterizing the alleged paradigm shift in Section 2. Sections 3 and 4 will then be devoted to discussing how that shift applies in the contexts of theory assessment and final theory claim respectively. Section 5 will synthesize those findings.

---

[1] The reasons for using the paradigm concept in the present context will become clearer later on. It should be emphasised, though, that calling the string theorists' perspective an emerging new paradigm does not per se prejudge that perspective's validity. Paradigms may well emerge at some stage just to be found deficient a little later.

2: The Role of Scientific Underdetermination

The classical paradigm of scientific theory assessment is based on one crucial principle: theory confirmation must rely on empirical data that reproduces predictions made by the theory in question. Scientific knowledge about the world consists of statements which have been empirically confirmed to a sufficient degree. Any substantial contribution to the extension of scientific knowledge about the world must be based on generating new empirically confirmed statements in one of two ways: either by generating new statements whose empirical implications match the existing data or by carrying out successful tests of so far unconfirmed theories.[2]

The described classical understanding of theory assessment emerged for good reasons. Empirical testing obviously plays a crucial role in distinguishing science from other forms of thinking about the world. The guiding role of empirical tests was essential for generating the impressive record of scientific success during recent centuries. Still, the strong focus put on empirical confirmation by the canonical understanding of scientific progress marginalizes an aspect of reasoning that arguably constitutes an important element of scientific thinking: the assessment of scientific underdetermination.

When a scientist develops a new theory in order to explain an observed phenomenon, her understanding of that theory's chances of being viable (i.e. predictively successful in the future) will depend on a number of considerations which reach beyond the theory's compatibility with the empirical data. For example, she might find the theory so simple and straightforward that she could not imagine it being false. She might be intrigued by the new theory's beauty and assume that such high level of beauty cannot go entirely wrong. She might also notice that the new theory can explain a number of previously unrelated phenomena and take that as an indication of the theory's viability. Finally, she might get the impression that there just is no other scientific solution to the problem addressed by her theory and infer that her theory is probably correct.

Looking at the cited arguments more carefully, one finds that the last argument is implicitly deployed in all others as well. A theory's simplicity, beauty or universality can only indicate its viability to an observer who assumes that no or very few alternative theories of comparable simplicity, beauty or universality can be fit to the available data. Assessments as to how likely it is that no or few alternative theories can be fit to the available data thus lie at the root of all considerations regarding the prospective viability of a so far empirically unconfirmed or insufficiently confirmed theory. We want to call such assessments 'assessments of scientific underdetermination'.

A few comments should be added in order to clarify the concept. First, it is important to note that conjecturing limitations to scientific underdetermination along the lines sketched above addresses a limited horizon of experimental testing. The scientist who resorts to this kind of reasoning wants to assess whether her theory can be hoped to get empirical confirmation of its core predictions within a certain experimental framework. Therefore, she is primarily interested in alternative theories which are empirically distinguishable at the next stages of empirical testing. Scientific underdetermination thus must address the expectations regarding the number of "bundles" of scientific theories which are compatible with the data D_t available at time t but are empirically distinguishable by a limited set E of future experiments which extend the tested regime along the directions

---

[2] Strict predictivists (see e.g. [Lakatos 1970]) have even insist that theory confirmation must be based on empirical data that has not entered the theory's creation process.

suggested by the given scientific context. A bundle of theories here is taken to consist of theories which may not be empirically equivalent with respect to all possible data but are indistinguishable by the set E of upcoming experiments. Underdetermination is significantly limited if only one or few such distinguishable bundles of theories can be constructed. We want to call this kind of reasoning a local assessment of scientific underdetermination. A more far-reaching 'global' assessment of scientific underdetermination would ask for the number of theories which fit a given set of available data and can be distinguished by the set of all data that could possibly be collected. Given the limited range of applicability of conventional scientific theories, global limitations to scientific underdetermination usually don't play a role in evaluating a theory's predictive prospects. We will come back to the distinction between local and global underdetermination in Section 4.

No observer of scientific praxis will deny that assessments of scientific underdetermination play a role in scientific reasoning. Scientists have opinions about the chances of success of new theories. In cases where they have already found many theories or models which are compatible with the given data but can be distinguished by upcoming experiments, they take each of those theories or models as a mere hypothetical example without good chances of being viable. In other contexts, scientists can be quite confident with regard to an empirically unconfirmed theory's viability. Scientists thus clearly make judgments on scientific underdetermination based on an analysis of the theory in question and on the context within which it was developed.

The canonical understanding of scientific progress, however, strictly distinguishes assessments of scientific underdetermination from the core elements of scientific progress, which are (1) the development of a scientific hypothesis and (2) the empirical testing of that hypothesis. The latter two steps are taken to generate scientific progress by creating new statements and establishing them as (probable) scientific knowledge. Assessments of scientific underdetermination, to the contrary, are taken to constitute mere instances of auxiliary reasoning that may be of some relevance by channeling scientific activity towards more promising investigations but do not directly contribute to the generation of scientific knowledge. Put in terms of the old conceptual dichotomy between context of discovery and context of justification, one may say that assessments of scientific underdetermination were always acknowledged as playing some role in the context of discovery but were denied any role in the context of justification.

It may be fair to say in hindsight that this understanding was never entirely satisfactory. If assessments of scientific underdetermination were not capable of raising the probability of a theory's viability, any reliance on them would be plainly irrational. Few, however, would want to assert that the scientists' judgments along those lines are completely senseless. If assessments of underdetermination are capable of raising the probability for a theory's viability, however, it seems difficult to deny that assessments of scientific underdetermination, if powerful enough, can amount to establishing (probable) scientific knowledge. A possible way out of this impasse is the understanding that assessments of underdetermination are not absolutely incapable of generating scientific knowledge but merely happened to be too weak for doing so up to now. Adopting that position regarding the scientific process up to the 1970s already brings us close to the claim that shall be made in the following.

We want to argue that, whatever the judgment regarding earlier phases of physics, assessments of underdetermination have attained a status in contemporary fundamental physics that makes them a crucial element of the acquisition of scientific knowledge and

cannot be neglected any more in a meaningful analysis of the mechanisms of scientific progress. The two contexts analyzed in the upcoming chapters will support this claim.

3: Trust in String Theory

Since the 1970s, theory building in particle physics has generated theories whose empirical viability could neither be empirically established nor refuted for several decades. One prominent example is supersymmetry, which may find empirical confirmation at the current LHC experiments at CERN - about four decades after the first formulations of supersymmetric models were developed. Another important example is the development of grand unified theories (GUTs), which were first formulated in the 1970s as well, but whose characteristic energy scale lies many orders of magnitude beyond the grasp of collider experiments.[3]

In some respects, string theory is even more detached from empirical confirmation than the aforementioned examples. Besides the remoteness of its expected characteristic scale, string theory faces an additional problem. Despite four decades of intense work, the theory has not even come close to being a complete theory. Though many of the theory's aspects have been analysed, no full understanding of the theory's structure has been reached or is in sight.

Moreover, string theory's predictive power is insufficiently understood on general grounds. The theory is presently believed to allow for a huge number of discrete local minima of the theory's potential (estimates go up to numbers of 10^{500} or 10^{1000}), the so called string theory landscape [Susskind 2003]. Since the selection of a ground state is a quantum process, it cannot be predicted based on the fundamental theory's structure; and since the characteristics of each groundstate translate into specific parameter values of low energy physics[4] a huge number of local ground states suggests that a correspondingly huge number of low energy patterns of parameter values are physically possible. Moreover, the impossibility to scan so many groundstates 'by hand' implies that it would be very difficult to understand whether the universe we live in corresponds to a valid string theory ground state even if the theory's fundamental structure was fully understood. In this light, it is difficult to assess how predictive string theory can be in principle.

A gradual alleviation of the empirical as well as the theoretical deficits of string physics in the future seems possible. At the theoretical level, the pace of difficult but consistent progress of string theoretical analysis during the last decades leaves hope that a gradually increasing amount of predictive conclusions from string physics might be forthcoming in the long run. Empirically, some string theoretical models (see footnote 1) suggest that stringy signatures might even be found at the LHC. If the string scale turned out to be as high as traditionally assumed and therefore excluded empirical confirmation in collider experiments, the current improvements of cosmological precision measurements in conjunction with the

---

[3] GUT theories predict proton decay, which could be measured without approaching the GUT scale in collider experiments but has not been discovered up to this point. The simplest GUT scenario (minimal SU(5)) is ruled out due to the lack of observed proton decay while in other scenarios the level of proton decay can be small enough for being compatible with the available empirical data. Proton decay can have other reasons than grand unification, which is why a conclusive confirmation of grand unification would be difficult to attain without measuring signatures of GUT-gauge bosons. The latter, however, must be expected to lie beyond the reach of any imaginable collider experiment.

[4] The term 'low energy physics' denotes physics that is testable at the current LHC experiment at CERN.

increasing interdependence between string theory and cosmology still could allow a certain level of empirical access to string theoretical predictions on that basis in upcoming years. Thus, one may hope that the connection between theory and experiment in string physics won't remain as inaccessible as it is today. There is no reason to expect, however, that string theory will either be fully understood theoretically or find conclusive empirical confirmation in the foreseeable future.

The highly influential position string theory has acquired since the mid 1980s stands in remarkable contrast to its theoretical incompleteness and its precarious empirical status. As already mentioned in the introduction, this high self confidence has in recent years led to criticism by a number of physicists who work in other fields or on other research programs.

We want to argue that the dispute on the status of string theory is in fact based on diverging understandings of the role of assessments of scientific underdetermination in science.[5] While the critics of string theory stick to the traditional understanding that assessments of scientific underdetermination cannot generate scientific knowledge, string theorists have been led by their scientific activity towards a perspective that acknowledges judgments based on assessments of underdetermination as legitimate means of establishing a scientific theory. This does not mean that the importance of empirical testing is denied. The new paradigm adheres to the understanding that conclusive empirical evidence which supports a theory's core predictions constitutes the strongest form of theory confirmation, trumps all other strategies of theory evaluation and therefore must be aimed at throughout the scientific process. The emerging new paradigm moves away from an understanding, however, that attributes the status of mere hypotheses to scientific theories which have found no empirical confirmation.

Of course, the drought of empirical evidence plays an important role in the described paradigm shift. It is crucial to understand, however, that it is not the only reason, arguably not even the most important one, for its occurrence. The main factor in the new paradigm's emergence is constituted by the fact that string theory itself as well as the context and dynamics of its evolution provide a workable and plausible basis for assessments of scientific underdetermination.

Three contextual arguments to that end will be discussed in the following. We will present a coherent reconstruction of strategies of reasoning which largely constitute informal background arguments that support the trust in string theory but tend not to be laid out explicitly in scientific works. Therefore, it is not easy to find classical formulations in the literature. All arguments to be discussed can be identified in a more or less explicit form in Polchinski's classical textbook [Polchinski 1998], however. It should be noted that the described arguments are by no means new inventions of string physics. They have played a role at earlier stages of the physical evolution but were largely ignored in the canonical understanding of theory assessment. This was possible for two reasons. First, empirical testing was available at that time and second, the arguments deployed were somewhat weaker than in contemporary fundamental physics. The changing circumstances from the 1980s onwards, however, render those arguments more conspicuous and increasingly difficult to ignore.

Only the first argument explicitly expresses an assessment of scientific underdetermination. However, a closer analysis of the other two arguments reveals their

---

[5] A more detailed analysis of the claim that criticism of string physics is due to a paradigmatic shift is given in [Dawid 2009].

reliance on limitations to underdetermination as well. The three arguments are interdependent and have to be combined to generate a powerful overall statement.[6]

(**1) The argument of no alternatives**: This argument constitutes an assessment of limitations to scientific underdetermination in a fairly straightforward way. String theorists tend to believe that their theory is the only viable option for constructing a unified theory of elementary particle interactions and gravity. It is important to understand the scope and the limits of that claim. String theory is not the only theory dealing with questions of quantum gravity. Various approaches to canonical quantum gravity try to reconcile gravitation with the elementary principles of quantum mechanics. The currently most popular among those approaches may be Loop quantum gravity [Rovelli & Smolin, 1990]. Those approaches (as well as some other, non-standard, approaches to a reconciliation of quantum physics and gravity) discuss the question of unification at an entirely different level than string theory, however. The latter stands in the tradition of the standard model of particle physics and is based on pivotal concepts such as non-abelian gauge theory, spontaneous symmetry breaking, and renormalizability. String theory's goal is to reconcile gravity with these advanced and successful concepts of contemporary particle physics and therefore to provide a truly unified description of all natural forces. At the present stage, the investigations in canonical quantum gravity, do not constitute alternatives at this level, which leaves string theory as the only available candidate theory.[7]

Specific support for the argument of no alternatives can be found at a number of levels. First, alternative attempts to achieve a consistent unification of nuclear interactions and gravitation have been tested quite extensively in the 1970s and 1980s in the context of supergravity theories. Today it is understood that supergravity cannot provide a satisfactory solution to the problems of non-renormalizability that arises in field theoretical attempts to carry out such unification. It can, however, be understood as an effective theory (a good approximation at lower energies) of string theory. Second, analysis within the framework of string physics could have but has not led to the emergence of alternative theories. Third, there are vague arguments that even attempts to sacrifice very basic physical principles in order to find alternative scenarios to string theory would, if made coherent, lead back towards the string theoretical approach (see e.g. [Polchinski 1999]).

The observation that the physics community cannot think of a viable alternative to string theory does not in itself imply that there are in fact no or few alternatives. It might just as well be an indicator of the limits of human intellect, creativity or scientific diligence. Moreover, candidates for alternatives do appear at times and thus pose a constant threat to simple arguments of no alternatives.[8] In order to establish that the lack of alternatives found

---

[6] An analysis of these arguments has been given in [Dawid 2007]. Bayesian formalizations of Arguments 1 and 2 are presented in [Dawid & Hartmann, 2011a and 2011b].

[7] That is not to deny the relevance of investigations in canonical quantum gravity. Many string theorists would just guess that, to the extent canonical quantum gravity finds physically viable solutions, those solutions will blend into the string theory research program once they are put into the context of contemporary high energy physics. No-one can exclude, of course, that ideas from canonical quantum gravity one day will develop into a full-fledged alternative theory of all interactions. If such an alternative theory was found, it would open up far-reaching new perspectives. If not, serious attempts to find it would have been significant nevertheless as contributions to the strengthening of the no alternatives argument.

[8] A recent example for an interesting new perspective is [Horava 2009], which proposes a scenario that solves the renormalizability problem of quantum gravity in a different way. It presents a theory that is non-relativistic at very short distances but approximates relativistic physics at longer distances. If consistent, such a scenario would constitute an alternative to string theory as a possible solution to the renormalizability problem of

so far does constitute a significant indication of a scarcity of possible options, it is necessary to consider arguments beyond the argument of no alternatives itself.

(2) The argument of unexpected explanatory interconnections closely resembles a well known principle of empirical theory confirmation. It is widely held that a truly convincing confirmation of a scientific theory must be based on those of the theory's achievements which had not been foreseen at the time of its construction. Normally, this refers to phenomenological predictions which are later confirmed by experiment. However, there is an alternative. Sometimes, the introduction of a new theoretical principle surprisingly provides a more coherent theoretical picture after the principle's theoretical implications have been more fully understood. This kind of theoretical corroboration plays an important role in the case of string theory. Once the basic postulate of string physics had been stated, a series of unexpected deeper explanations of seemingly unconnected facts or theoretical concepts emerged. Two strong examples for this development are the interpretation of black hole entropy and the connections to supersymmetry and supergravity.

Initially, interest in supersymmetry was motivated primarily by the abstract mathematical question whether any generalisation of the classical symmetry groups was possible. As it turns out, supersymmetry is the maximal consistent solution in this respect. Soon after the construction of the first supersymmetric toy-model, it became clear that a formulation of supersymmetry as a gauge-symmetry (=local supersymmetry or supergravity) had the potential to provide a fuller understanding of the particle character of gravity. In the context of string theory, on the other hand, it had been realised early on that a string theory that involves fermions must necessarily be locally supersymmetric.[9] The question of the maximal symmetry group, the quest to integrate the graviton naturally into the field theoretical particle structure, and the attempts to formulate a consistent theory of extended elementary objects thus miraculously blend into one coherent whole.

A problem that arises when general relativity goes quantum is black hole entropy. The necessity to attribute an entropy proportional to the area of its event horizon to the black hole in order to preserve the global viability of the laws of thermodynamics was already understood in the 1970s. The area law of black hole entropy was merely an *ad hoc* posit, however, lacking any deeper structural understanding. In the 1990s it turned out that some special cases of supersymmetric black holes allow for a string theoretical description where the black hole entropy can be understood in terms of the number of degrees of freedom of the string theoretical system [Strominger & Vafa 1996]. Thus, string physics provides a micro-physical understanding of black hole entropy.

All of these explanations represent the extendedness of particles as a feature that seems intricately linked with the phenomenon of gravity and much more adequate than the idea of point-particles for a coherent overall understanding of the interface between gravity and microscopic interactions. The subtle coherence of the implications of the extendedness of elementary objects could not have been foreseen at the time when the principle was first suggested. It would look like a miracle if all these instances of delicate coherence arose in the context of a principle that was entirely misguided.

The argument of unexpected explanatory interconnections can be read as an argument against a high degree of scientific underdetermination. If many solutions to the

---

quantum gravity. The debate on the scenario's consistency, its promises and its limitations is ongoing at this point.
[9] World sheet supersymmetry of a string that includes fermions was discovered by [Gervais & Sarita 1971]. A string theory that shows local target space supersymmetry was finally formulated by [Green & Schwarz 1984].

problem of providing a finite or renormalizable theory of quantum gravity existed, there would be no reason to expect that most of them provide the unexpected explanatory interconnections provided by string physics. If only one solution existed, on the other hand, that solution must provide these explanatory interconnections because a fully valid theory of quantum gravity must account for all aspects of the phenomena involved. Observing these interconnections thus may be taken as an indicator that only one or few solutions exist.

While the argument of unexpected interconnections thus can support an argument of no alternatives, it cannot be conclusive. It may still be the case that, for some unappreciated deep structural reason, all solutions to one problem have to be solutions to a set of other problems as well. A structural reason of that kind would explain unexpected explanatory interconnections. It would not exclude, though, that a very large number of theories on that basis can explain the available data. Therefore, the argument of unexpected internal coherence on its own does not strictly imply limitations to scientific underdetermination.

Based on arguments 1 and 2, we can draw the following preliminary conclusion. The two arguments seem to present a coherent case for limitiations of scientific underdetermination, which, if true would translate into good chances that string theory is viable. However, an alternative interpretation is possible. If the lack of known alternatives is due to the scientists' limited capabilities and the unexpected explanatory interconnections are due to some deeper structural reasons which do not enforce a specific theory, arguments 1 and 2 do not work as reliable indicators for string theory's viability. In order to understand how likely this alternative interpretation is, we have to resort to a third argument, which connects our line of reasoning to empirical data.

(3) The meta-inductive inference from other cases of predictive success. In string theory, this argument strongly relies on the example of the standard model of particle physics [Weinberg 1967]. The standard model was developed in the 1960s and early 1970s in order to provide a coherent description of electromagnetic and weak interaction processes that could account for the observed fermionic masses. In the early 1970s, it turned out that the concept of non-abelian gauge field theory deployed in the standard model was a promising tool also for describing strong interaction: the properties of confinement and asymptotic freedom observed in the context of strong interactions could be shown to be directly implied by unbroken non-abelian gauge field theory. The resulting application of non-abelian gauge theory to all nuclear interactions resulted in the full formulation of the standard model as we know it today. Initially, the use of non-abelian gauge field theories was largely motivated by the aim to solve a serious consistency problem of particle physics: the non-renormalizable character of simpler models of weak interaction. Starting from there, gauge field theory turned out to provide a good solution to the problems of confinement and asymptotic freedom in strong interactions (thus providing a good example of unexpected explanatory interconnections). After the renormalizability of non-abelian gauge theory had been proved in the early 1970s ['t Hooft & Veltman 1972], the standard model was the only known satisfactory theory of all nuclear interactions (thereby providing the basis for an argument of no alternatives).

The standard model was based on the conjecture of a new class of non-abelian gauge vector fields which had not been empirically confirmed at the time of the theory's construction. Furthermore, it predicted a wide spectrum of so far unconfirmed fermionic particles and symmetries of the interaction structure. In the two decades until 1994, all those predictions except for one (the Higgs particle that is presently searched for at the LHC experiments in Geneva) got empirically confirmed. The standard model of particle physics

today thus constitutes a strong example of a theory that (i) did not have any equally satisfactory alternatives before its core characteristics had been empirically confirmed, (ii) revealed significant unexpected explanatory interconnections after its construction and (iii) indeed turned out to be empirically viable at the end. Given its eventual empirical success, the arguments of no alternatives and of unexpected explanatory interconnections appear as reliable indicators of the theory's viability.

The case of the standard model fits into a wider pattern. There are a number of striking cases in 20th century physics where an explanatory powerful theoretical solution for a crucial consistency problem seemed without equivalent alternatives and eventually turned out to be predictively successful. In microphysics, one may cite the successful prediction of anti-matter based on Dirac's formulation of relativistic quantum theory and subsequent correct quantitative predictions in the framework of quantum electrodynamics. In gravitational physics, the most striking example would be the successful prediction of light bending by Einstein's general theory of relativity. On the other hand, there seem to be few if any recent examples in fundamental physics where conditions (i) and/or (ii) from above were clearly fulfilled but the corresponding theory turned out predictively unsuccessful when empirically tested. This wider framework may suggest that arguments (1) and (2) can serve as reliable indicators for a theory's viability in the context of fundamental physics.

The argument becomes more cogent when expressed in terms of scientific underdetermination. Like arguments (1) and (2), the standard model's predictive success indicates limitations to scientific underdetermination: if large numbers of alternative theories with different empirical predictions existed, physicists could have found many theoretical solutions which differed from the standard model and were empirically inadequate. It would then seem miraculous that the predictively successful theory and none of its alternatives was found by the scientists in the 1960s. The assumption of a very small number of possible solutions, to the contrary, can explain why the scientists found an empirically successful solution.

We thus have a plausible understanding why (iii) predictive success tends to occur in conjunction with (i) a scarcity of known alternatives and (ii) unexpected explanatory interconnections: all three phenomena are natural consequences of limitations to scientific underdetermination. If we observe that, within our research field, (i) and (ii) indeed tend to occur in conjunction with (iii) whenever sufficient empirical testing has been carried out, this justifies the conclusion that we work within a research field where possible alternative reasons for (i) and (ii) (which have been mentioned at the end of the discussion of argument (2)) tend not to apply: scientists tend to be capable enough for finding alternative theories if such alternatives abound, and unexpected explanatory interconnections tend not to arise due to deep structural reasons which allow for many alternative theoretical realizations. It therefore seems reasonable to carry out a meta-inductive inference from the predictive success of theories like the standard model to the probable viability of other so far empirically unconfirmed theories where, just like in the standard model case, (i) and (ii) are realized.

Note that argument (3), just like arguments (1) and (2), would not be convincing on its own. Looking at argument (3) alone, it would seem unclear whether the predictive success of theories like the standard model or general relativity can be relevant for an assessment of the viability of string theory. After all, the latter looks quite different from the

former theories in a number of respects[10]. One has to look at arguments (1) and (2) in order to understand in which respects all theories under consideration are similar and why that similarity is important when addressing the question of a theory's viability. Only the conjunction of all three non-empirical arguments of theory assessment therefore can instil substantial trust in the viability of string theory even in the absence of empirical confirmation.

4: The Final Theory Claim

String theorists tend to assume that a coherent formulation of string theory would constitute a final theory: no other more fundamental or more universal theory than string theory would ever be required to account for new empirical data. This form of a final theory claim is a little weaker than some final theory claims discussed elsewhere in the literature. First, it is confined to physics and does not address the question of a unified description of physics and mentality (which is at the core of the analysis in [Chalmers 1996]). Moreover, its empirical focus makes it weaker than final theory concepts which are based on the notion of ultimate explanation (see [Rescher 2006]).[11]

The role played by final theory claims in string physics is reminiscent of the role played by non-empirical assessments of the theory's viability. Final theory claims clearly contribute to the fascination of string theory and arguably constitute a significant element of the overall mindset behind string physics. However, they seem to lie a little beyond the concerns of down to earth science, which is why their analysis mostly remains at an informal level and is more conspicuous in private communication and in publications aimed at a wider public than in scientific publications. The final theory claim was most famously discussed in Stephen Weinberg's book "Dreams of a Final Theory" [Weinberg 1994], it is addressed in [Greene 1999] and in [Kaku 1997] and has been emphasised at several instances by S. Hawking (see e.g. [Hawking & Mlodinov 2010]).

String theorists have two main reasons for adopting the belief that string theory might be a final theory. First, string theory unifies nuclear interactions and gravity and therefore, if valid, would cover all physical interactions known today. If full universality is taken as a goal of physical theory building, string theory thus looks like a candidate for a theory that has achieved this goal and therefore may be taken to terminate the sequence of theories moving in that direction. Second, theory succession in fundamental microphysics throughout the 20th century has been driven by moving towards testing smaller and smaller distance scales. String theory turns out to predict that its own characteristic scale, the string length, constitutes a fundamental minimal distance scale. In a string theoretical context, all information that might be framed as information on smaller distance scales than the string length is redundant and can be expressed also in terms of information on scales higher than the string scale. The crucial theoretical concept behind this statement is T-duality, which roughly provides a transformation that translates all statements about distance scales R into

---

[10] Though being developed as an attempt to solve a consistency problem just like string theory, the standard model obviously was much closer to empirical data as string theory and immediately generated very specific empirical predictions. General relativity, whose development was more similar to the one of string theory in being motivated by very general reasoning without close interaction with empirical data, was developed within a significantly different conceptual framework.

[11] The conceptual problems identified by Rescher with respect to his definition of a final theory claim thus do not arise in our case.

empirically equivalent statements about distance scales α^2/R (where α is the string length). (For a discussion of T-duality' implications for a minimal length scale, see [Witten 1996]

Critics of string theory often consider final theory claims unscientific. The general sentiment behind critical remarks published in some books and articles (see e.g. [Dyson 2008]), is arguably shared by many physicists and philosophers of science. The most straightforward arguments against final theory claims work at an epistemic level: even if one granted the theoretical possibility of a final theory, it seems that no sound epistemic basis can justify attributing that status to any specific theory. Regarding the universality argument, the following problem arises. Statements on a theory's full universality can only be made with respect to the set of phenomena known at the time. Such statements do not imply, however, that new phenomena which reach beyond the allegedly universal theory cannot be discovered in the future. Equally, the duality argument for string theory's final character can be criticized on general grounds. Whatever reasons a theory might offer for being seen as a final theory, those reasons must always be based on the theory itself. Thus, they can never constitute conclusive arguments against the possibility that an altogether different theory could supersede the present one and render the theory-dependent final theory claim obsolete. Any final theory claim in this light seems to be based on an unfounded denial of the historicity of scientific reasoning.

In the following, we will argue that the rising importance of assessments of scientific underdetermination devaluates the epistemic arguments against final theory claims and thus gives new credence to the latter. In fact, both stated arguments against final theory claims are based on the canonical paradigm of theory assessment according to which a theory's status can only be assessed to the extent it has been tested by empirical data. Since the available data will never exhaust the data that could be collected in principle, judgements regarding the status of scientific theories can never address the question of finality.

It has already been argued in the previous section that the canonical paradigm of theory assessment does not square with the actual research process in contemporary high energy physics. Assessments of limitations to scientific underdetermination reach beyond the horizon set by currently available empirical data. Can those assessments be extended towards a strengthening of final theory claims?

To answer this question, we first have to understand that the case of final theory claims differs substantially from the case of non-empirical theory assessment discussed in Section 3. Section 3 has established the existence of rational arguments for what we have called 'local' limitations to scientific underdetermination. A scarcity of alternatives was argued for with respect to bundles of theories which give distinguishable predictions regarding some limited set of future experiments. A final theory claim, to the contrary, must be based on an assessment of 'global' limitations to underdetermination, that is, limitations to the number of theories which are empirically distinguishable based on all possible observations. If we want to use the arguments presented in Section 3 for supporting final theory claims, we therefore have to establish that, under appropriate circumstances, assessments of local limitations to scientific underdetermination can support the case for global limitations as well.

Usually, local limitations to underdetermination can be established without implying any global limitations. In special sciences, this fact arises as a consequence of what one could call the vertical hierarchy among scientific theories. A clear distinction can be drawn between the characteristic length scale of some phenomenon to be explained by a theory in a special science and smaller length scales that characterise more fundamental scientific

theories on which the theory in question is implicitly based. Thus, even if only few mechanisms can explain the available data and provide distinguishable predictions for future experiments at one level of description (which means that local underdetermination is limited at this level), many different realisations of those mechanisms may be possible at the micro-levels addressed by more fundamental theories.

To give an example, any modern biological theory about the evolutionary mechanisms that led to the development of wings is implicitly based on the understanding - drawn from substantially different fields of scientific research - that animal wings consist of atoms which, in turn, are built of more fundamental elementary particles. Evolutionary biologists now may argue for the viability of an individual theory about the evolutionary development of wings in a specified context by pointing out that the theory in question seems to be the only theory that can account for all known empirical evidence relevant in the given context. (That is, they deploy a claim of local limitations to scientific underdetermination.) This does not imply, however, that only one kind of microphysical realisation or only one set of bio-chemical or physical laws could underlie the conjectured macroscopic evolutionary mechanism. Which microphysical realization actually applies is not subject of the initial investigation and has to be established through experiments at entirely different energy scales carried out in biochemistry and microphysics. Local limitations to scientific underdetermination thus do not translate into global ones in the given case.

In fundamental physics, the distinction between local and global underdetermination works in a similar way as in the special sciences. The role played by the hierarchy of theories in the special sciences is now taken over by the limited range of viability of individual theories within a more universal general conception of physical phenomena. We may know about a theory B which constitutes an alternative to a theory A but which is characterized by a certain parameter that has been measured already within a physical context unrelated to theory A. Let us now assume that the measured parameter value lies far beyond the range of the next generation of experiments E which have been planned in order to test theory A. In that case, we know that theory B, while relevant for our understanding of global underdetermination, remains irrelevant for local underdetermination: theory B does not constitute an alternative theory that, according to our best knowledge, can be distinguishable from A by experiments E.

To give an example for this way of reasoning, let us look at non-relativistic quantum mechanics. Quantum field theory constitutes a theoretical alternative to non-relativistic quantum mechanics that becomes empirically distinguishable from the latter once processes close to the velocity of light enter the picture. Physicists at the time when quantum physics was developed knew the speed of light from astronomical measurements which had no connection to quantum physics. That knowledge fixed the scale where non-relativistic quantum mechanics and quantum field theory became distinguishable and, on that basis, created a regime where quantum field theory was known not to constitute an empirically distinguishable alternative to quantum mechanics. Therefore, quantum field theory left local limitations to underdetermination within that regime untouched and only contributed to global underdetermination.

Having established the nature of the distinction between local and global underdetermination in fundamental physics, we now come to the crucial point of our analysis. As we will show, the final theory arguments presented above can be understood as arguments which block the distinction between local and global underdetermination. To understand this, let us put the relevant limit for local underdetermination at the characteristic scale of our theory where the theory's core predictions can be tested. (In the

case of string theory, it would obviously make no sense to discuss local underdetermination with respect to the next generation of collider experiments which most likely are incapable of testing the theory.) Now, let us first look at the case of a fundamental theory that is fully universal in the sense that it covers all information on parameter values which determine the size of known phenomena. In this case, no external theory is available anymore that could fix a parameter value that controls the extent to which our theory and a possible alternative theory are empirically distinguishable. The kind of reasoning sketched in the previous paragraphs therefore breaks down.

This step already significantly reduces the options for alternative theories which do not affect local limitations to scientific underdetermination. It does not fully eradicate them, however. We could still face situations where a certain constellation within the present theory constrains alternative theories to regimes which remain inaccessible to those experiments which test the present theory's core predictions. An example for such a constellation would be the relation between the standard model and grand unified theories. Grand unified theories are compatible with present empirical data only if unification occurs at the scale where the known gauge couplings (which have been measured close to the electroweak scale and can be calculated for higher energy scales based on renormalization group techniques) all assume the same value. This scale, called the GUT scale, is known to lie far beyond the grasp of collider experiments, however (if we want to disregard the possibility of large extra dimensions for the moment). On that basis, we can know that grand unified theories cannot be distinguished from the standard model in collider experiments which test energy scales far below the GUT scale. Grand unified theories thus constitute alternatives which affect global underdetermination but most probably are of no concern to local underdetermination immediately above the electroweak scale.

Constellations of the kind described in the last paragraph, however, are excluded by the second final theory argument, the argument for a minimal length scale based on duality. Once such a minimal scale is established within a universal theory at that theory's characteristic scale, there is no way to make sense of the statement that this theory remains viable beyond its characteristic scale but stops being viable far beyond that scale. Expressed in terms of energy scale, any new theory that is distinguishable from the old theory at all thus must be distinguishable already one step beyond the scale where the old theory's core predictions are empirically tested. In other words, if one would find an alternative theory that contributed to global scientific underdetermination it necessarily would have to contribute to local underdetermination as well.[12]

Both final theory claims presented in the context of string theory in this light may be understood in a new way. On their own, the arguments cannot establish that theory succession which goes beyond string theory is unlikely. They do establish, however, that arguments which suggest limitations to local scientific underdetermination must be acknowledged as arguments against unlimited global scientific underdetermination as well.

---

[12] Note that the concepts of local and global underdetermination have been defined with respect to the energy scale, which, as discussed above, defines the range of experimental testing in high energy physics. Other parameters could control a theory's empirical testability as well, though. Examples would be the number of events or the time scale of empirical testing. Those parameters may be theoretically related to the theory's characteristic energy scale. (A good example is proton decay, where the phenomenon's observation depends on the number of protons observed but the theoretical framework relates proton decay to new physics at some high energy scale.) If no such theory-based correlation with energy scales exists, however, new physics might arise without having any characteristic energy scale. The presented argument does not exclude or render less likely new physics of that kind. An assessment of the likelihood of such scenarios must be based on different considerations.

The arguments thus play the role of mediators which connect the local and the global level of analysis and thereby raise the significance of local arguments.

Once the local limitations to scientific underdetermination in contemporary high energy physics have been shown to translate into global limitations, the viability of internal final theory claims can be related to empirical data via the meta-inductive argument and thereby can attain a certain degree of trust.

Acknowledging that final theory arguments can indeed be meaningful, it is important to specify the implications a trustworthy final theory claim can have in the context of contemporary string physics. Given the actual situation in string physics, such a claim clearly does not suggest anything like an imminent completion of theory building in fundamental physics. Rather, it seems to suggest a shift of the overall dynamics of the research process. The classical picture of scientific progress was based on the concept of an infinite sequence of scientific theories which attained a largely complete form within a limited period of time and could be tested empirically afterwards. The emerging new picture does not alter the infinite time horizon for the completion of physics. However, that infinite time horizon is now being identified with the time horizon for the completion of the one fully universal theory physicists are working on. In the old framework, we could not expect that our limited theories could ever cover the whole range of all empirical data that could ever be collected. In the newly emerging framework, the universal theory, call it string theory, can neither be expected to find a complete form nor conclusive empirical testing in the foreseeable future.

5: Conclusion

The previous sections have demonstrated that assessments of limitations to scientific underdetermination are of crucial importance at two different levels in string physics. They are essential for non-empirical assessments of a theory's viability: without establishing local limitations to scientific underdetermination, string physicists would have to treat their theory as a mere speculation and could not find any rational reasons for having any trust in its viability. Furthermore, assessments of limitations to scientific underdetermination also provide arguments for the theory's final theory status. Final theory claims amount to extending the local claims of limitations to scientific underdetermination to a global level.

As we have seen, the local and the global level of analysing scientific underdetermination show a complex pattern of mutual interdependence. Several layers of argumentation are built on top of each other. The basis for all other arguments of limitations to scientific underdetermination is provided by a use of empirical confirmation of scientific theories at a meta-level. The predictive success of scientific theories in the research field constitutes a necessary precondition for the viability of any other argument on scientific underdetermination by establishing the connection between entirely theoretical assessments and the performance of theories in an empirical context. Once this connection has been established, other purely theoretical arguments can be deployed to strengthen the case for local limitations to scientific underdetermination and thereby enhance the trust in a physical theory even in the absence of sufficient empirical confirmation. Eventually, global arguments on limitations to scientific underdetermination can be supported by the empirical basis as well. This last step works with regard to fully universal theories where local arguments of limitations to scientific underdetermination turn out to be relevant for the

question of global scientific underdetermination. On that basis, final theory claims may be turned into genuinely scientific statements.

The suggested mechanism constitutes a significant modification of the canonical paradigm of theory assessment. The assessment of scientific underdetermination is being acknowledged as a possible basis for the acquisition of scientific knowledge even in the absence of confirming empirical evidence for the corresponding theory. We have pointed out that the current developments in fundamental physics offer two good reasons for acknowledging assessments of limitations to scientific underdetermination as legitimate strategies of scientific theory evaluation. First, the increasing difficulties to obtain conclusive empirical confirmation may be taken to suggest that fundamental physics is must resort to these strategies in order to remain a sound and dynamical scientific field. Second, our analysis has provided strong indications that the new strategies of theory assessment can indeed be justified. They are conceptually related to and based on the principle of empirical testing and therefore do not constitute arbitrary criteria of scientific success.

In order to acknowledge the deployment of the new strategies as a critical method of theory assessment, it is important to consider one more question, though: Is it possible to construe scenarios of non-empirical evidence which would weaken or destroy trust in these strategies in a given context and consequently reduce trust in the theories supported by them? Indeed, such scenarios can be found for all presented arguments of non-empirical theory assessment. (For a slightly more extensive analysis of this point see [Dawid 2009].)

The argument of no alternatives relies on the observation that no conceptually equally satisfactory alternatives to string theory are found. It would have to be withdrawn whenever a conceptually satisfactory alternative to string theory was in fact discovered. Any such alternative would thus significantly reduce the trust in string theory. The argument of unexpected explanatory interconnections would lose strength if some of the interconnections in question turned out to be based on more general patterns of reasoning which are not univocally related to string theory.[13] On a more general basis, theoretical arguments in favour of string theory could also be weakened by other developments which would cast doubt on the research program's chances of success. For example, an improved theoretical understanding of string physics might reveal theoretical weaknesses which change the theoretical assessment of the theory's chances of being fully consistent and thus physically viable. Furthermore, an interruption of theoretical progress over a long period of time could raise doubts whether a more complete theoretical understanding of string physics was attainable at all.

The meta-inductive argument from other theories' predictive successes, finally, can lose power based on new empirical data regarding other theories. Trust in string theory could be substantially reduced if theories which seemed to have no alternatives turned out to make false predictions. To give an example, let us assume for a moment that no Higgs particle was found at the ongoing LHC experiments at CERN. Many observers would then raise the question how confident one could be regarding the remote claims of string theory if one could not even predict the existence the Higgs particle correctly.

Regarding final theory claims of string physics, any argument that would weaken the status of string theory itself would weaken the related final theory claim as well.

---

[13] In fact, considerations along these lines are being discussed and contribute to a better understanding of the significance of the argument of unexpected explanatory interconnections. For example, it has been argued that microscopic calculations of black hole entropy can be calculated based on general principles without using a string theoretical framework (see [Strominger 1998], [Carlip 2008]).

String theory thus could well fail according to non-empirical theory assessment even in the absence of empirical refutation. Non-empirical theory assessment therefore can be called a critical method of theory evaluation. At a meta-level, the non-empirical strategies of theory assessment are themselves strengthened or weakened based on comparing theory assessments on their basis with later results of empirical testing. Any instance where a theory's endorsement based on non-empirical theory assessment is later vindicated by empirical testing strengthens the principle of non-empirical theory assessment; any instance where such an endorsement must be withdrawn due to empirical testing or new theoretical analysis weakens it. The future relevance of the discussed strategies of non-empirical theory assessment therefore will strongly depend on the future scientific evolution.

The fact that the reliability of the method of non-empirical theory assessment can itself be empirically tested provides a basis for taking seriously judgements like the final theory claim, which cannot be confirmed empirically themselves but can be made plausible by being integrated into an empirically testable context.

Non-empirical theory assessment thus crucially relies on empirical testing and can never fully replace it. Nor does non-empirical theory assessment award the same status to a theory as strong empirical confirmation. It is vaguer and less conclusive than the testing of theories by empirical data. Its vagueness induces the risk that its deployment might be overstretched and thereby lose its cogent connection to the principle of empirical testing. The constant surveillance of strategies of non-empirical theory assessment thus seems of high importance.

If handled with care, however, these strategies can play an independent role in contexts where empirical testing is absent or insufficient for a long period of time. They may also reach beyond the limits of what can be empirically tested in principle to the extent the method of reasoning remains empirically testable. Non-empirical theory assessment breaks new ground in replacing the old dichotomy between empirical confirmation and mere speculation by a continuum of degrees of credibility, where the available elements of empirical corroboration and non-empirical theory assessment jointly contribute to an overall evaluation of theory's chances of being viable.

The present paper has been concerned with the role of assessments of scientific underdetermination in string physics. It is important to point out, however, that string physics merely constitutes a particularly strong example of a broader development. It has been mentioned above that theories like supersymmetry and grand unified theories have been developed without direct empirical confirmation for several decades just like string theory. Their chances of being viable have been assessed based on the very same local non-empirical arguments which are being deployed in the context of string physics. Empirical cosmology enters a phase where theory building becomes more specific and empirical data is getting more powerful, so that some degree of empirical testing can be established. Still, those empirical tests mostly are not strong enough for providing conclusive verdicts on specific cosmological theories. Assessments of scientific underdetermination in that case are joined with empirical data and assume an important role for an overall assessment of the theories' viability.

Many physicists may wish back the golden old days of physics when fundamental theories could (more often than not) be tested empirically within a reasonable period of time and a clear-cut empirical verdict in due time rendered irrelevant all tedious theoretical considerations concerning a theory's viability. Empirical science, however, must answer to the situation it actually faces and make the best of it. A sober look at the current situation in

fundamental physics suggests that the old paradigm of theory assessment has lost much of its power and new strategies are already stepping in.

References

Carlip, S. 2008: 'Black Hole Entropy and the Problem of Universality', arXiv:0807.4192.

Chalmers, D. 1996: *The Conscious Mind: In Search of a Fundamental Theory*, Oxford University Press.

Dawid, R. 2007: 'Underdetermination and Theory Succession from the Perspective of String Theory', Philosophy of Science, 73/3, 298-322.

Dawid, R. 2009: 'On the Conflicting Assessments of the Current Status of String Theory', Philosophy of Science,76/5, 984-996.

Dawid, R. And Hartmann, S. 2011a: 'No Alternative Arguments', forthcoming.

Dawid, R. And Hartmann, S. 2011b: 'No Alternative Arguments and Meta-Inductive Reasoning in String Theory', forthcoming.

Dyson, F. 2008: 'The Scientist as a Rebel', New York Review Books.

Gervais, J. L. & Sakita, B. 1971: 'Field Theory Interpretation of Supergauges in Dual Models', Nuclear Physics B34, p632.

Green, M. B. & Schwarz, J. H. 1984: 'Anomaly Cancellation in Supersymmetric D=10 Gauge Theory and Superstring Theory', Physics Letters 149B, p117.

Greene, B. 1999: *The Elegant Universe*, W. W. Norton.

Hawking, S. and Mlodinov, L. 2010: *Grand Design*, Bantam Press.

Horava, P. 2009: ,Quantum Gravity at a Lifshitz Point', Physics Review D79:084008; arXiv:0901.3775.

Kaku, M. 1997: *Beyond Einstein: Superstrings and the Quest for the Final Theory*, Oxford University Press.

Lakatos, I. 1970: ,Falsification and the Methodology of Scientific Research Programs', in Lakatos, I and Musgrave, A (eds), *Criticismand the Growth of Knowledge*, Cambridge University Press.

Polchinski, J. 1998: *String Theory*, 2 vol., Cambridge University Press.

Polchinski, J. 1999: 'Quantum Gravity at the Planck Length', hep-th/9812104, International Journal of Modern Physics. A14, 2633.

Penrose, R. 2005: *The Road to Reality,* Vintage Books London.

Rescher, N. 2000: 'The Price of an Ultimate Theory', Philosophia Naturalis 37, 1-20.

Rovelli, C. And Smolin, L., 1990: 'Loop Space Representation of Quantum General Relativity', Nuclear Physics B331, 80.

Smolin, L. 2006: *The Trouble with Physics*, Houghton Mifflin.

Strominger, A. 1998: 'Black hole Entropy from Near Horizon Microstates', JHEP 9802, 009; hep-th/9712251.

Strominger, A. & Vafa, C. 1996: 'Microscopic Origin of the Bekenstein-Hawking Entropy', hep-th/9601029, Physics Letters B379, p99.

Susskind, L. 2003: 'The Anthropic Landscape of String Theory', in Carr, B. (ed.), *Universe or Multiverse?*, Cambridge University Press, 247-266; hep-th/0302219.

't Hooft, G. and Veltman, M.J.G. 1972: 'Regularization and Renormalization of Gauge Fields', Nuclear Physics B44, 189-213.

Weinberg, S. 1967: 'A Model of Leptons', Physical Review Letters 19, 1264-1266.

Weinberg, S. 1994: *Dreams of a Final Theory,* Vintage.

Witten, E. 1996: 'Reflections on the Fate of Spacetime', reprinted in Callender, C. & Huggett, N., *Physics meets Philosophy at the Planck Scale*, Cambridge University Press 2001.

Woit, P. 2006: *Not Even Wrong: The Failure of String Theory and the Continuing Challenge to Unify the Laws of Physics*, Jonathan Cape.