Challenging Scientific Methodology: Theory Assessment and Development in Modern Fundamental Physics

Inaugural-Dissertation

zur Erlangung des Doktorgrades der Philosophie

an der Ludwig-Maximilians-Universität

München

vorgelegt von

Radin Dardashti

aus

Tehran, Iran

München 2016

Erstgutachter: Prof. Dr. Stephan Hartmann

Zweitgutachter: Priv.-Doz. Dr. Richard Dawid

Tag der mündlichen Prüfung: 4.07.2016

"Aus kleinen Missverständnissen gegenüber der Wirklichkeit zimmern wir uns Glaubensvorstellungen und Hoffnungen zurecht und leben von den Brotrinden, die wir Kuchen nennen, wie arme Kinder, die Glücklichsein spielen."

Fernando Pessoa

Contents

List of Figures

Für meine Eltern

Zusammenfassung

Theorien wie die Supersymmetrie, das Inflations-Modell und Theorien der Quantengravitation werden nun seit mehr als drei Jahrzehnten von Wissenschaftlern verteidigt, obwohl es an empirischen Belegen fehlt. Diese veränderte Situation, in der sich die Grundlagenphysik heutzutage befindet, macht es notwendig alternative Methoden der Theorienbestätigung in Betracht zu ziehen. Zum einem, um Theorien in Abwesenheit von empirischen Belegen zu testen und zum anderen, um mögliche Erklärungen zu finden, wieso Wissenschaftler ihren Theorien ein so großes Vertrauen schenken. Der hier vertretene Ansatz baut auf dem Konzept des Theorienraums, d.h. dem Raum aller wissenschaftlicher Theorien, auf. Die Idee ist folgende: der Standard für die Theorienbestätigung waren und sind weiterhin empirische Belege, da dies die verlässlichste Methode darstellt. Sobald wir jedoch empirische Tests als eine Einschränkung des Theorienraums betrachten, öffnen sich für uns neue Wege, die zur Bestätigung einer Theorie führen könnten. Im Prinzip können Belege jeglicher Art, sobald sie den Theorienraum einschränken als mögliche Strategie zur Theorienbestätigung in Betracht gezogen werden. Zudem lässt sich feststellen, dass der eingeschränkte Theorienraum hilfreiche Informationen im Hinblick auf die Theorienentwicklung liefert.

In der vorliegenden Dissertation wurden einige Zusammenhänge zwischen Theorienbeurteilung (in den meisten Fällen handelt es sich hierbei um die Theorienbestätigung im Bayesianischen Sinne) und Theorienentwicklung sowie des Konzepts des Theorienraums behandelt. Im

Folgendem fassen wir die Ergebnisse zusammen:

Das Konzept der Einschränkung der wissenschaftlichen Unterbestimmtheit stellt die Grundlage für die nicht-empirische Theorienbestätigung dar. Die von Richard Dawid vorgeschlagenen Argumente für die Einschränkungen der wissenschaftlichen Unterbestimmtheit sind indirekte Methoden, um den Theorienraum zu beurteilen. Indirekt in dem Sinne, dass nicht der Theorienraum selbst untersucht wird, sondern Belege die Einblicke in den Theorienraum erlauben. In Kapitel 2 betrachte ich die Möglichkeit den Theorienraum an sich zu bewerten. Dies führt zu drei Problemen.

Das Theorie-Problem besagt, dass Wissenschaftler bei der Theorienentwicklung zwei theoretische Fehler machen können bei denen der Theorienraum unbegründet eingeschränkt wird. Zum einen werden theoretische Annahmen eingeführt (wie die Annahme, dass jede Theorie Lorentz-invariant sein muss) und zum anderen kann das zu lösende Problem selbst theoretischer Natur sein. Beides ist problematisch, da diese Einschränkungen nicht ausschließlich auf die empirisch bestätigten Aspekte der Theorie basieren. Das Struktur-Problem bezieht sich auf das Problem, dass Wissenschaftler sich bei der Formulierung ihrer Theorien auf bestimmte mathematische Strukturen beziehen müssen, in denen die physikalischen Annahmen realisiert sind. Diese Strukturen sind jedoch meistens nicht eindeutig und können somit irrtümlicherweise den Theorienraum einschränken. Das letzte Problem stellt das Daten-Problem dar. Hierbei gehen Wissenschaftler fälschlich davon aus, dass gewisse empirische Aussagen (wie z.B. die dreidimensionalität des Raumes) ohne Einschränkungen gelten müssen. So wird erneut der Theorienraum vorzeitig eingeschränkt. Diese Probleme führen zu einer notwendigen Reorientierung in der wissenschaftlichen Praxis, falls Theorien mit Hilfe einer Beurteilung des Theorienraums bestätigt werden sollen.

Nachdem wir uns mit den Problemen der Theorienraumbewertung befasst haben, wenden wir uns in Kapitel 3 den vorgeschlagenen Argumenten Dawids zu. Insbesondere wenden wir uns dem Keine-Alternativen Argument (NAA) und der damit einhergehenden Bestätigung zu. Wir zeigen, inwiefern der verwendete nicht-empirische Beleg des NAA unzureichend ist und schlagen eine adäquatere Form vor. Diese erlaubt es das NAA auf Theorien der Quantengravitation anzuwenden. Wir argumentieren, dass die vorhandenen Einschränkungen auf den Theorienraum nicht ausreichen um diese Theorien zum jetzigen Zeitpunkt zu bestätigen.

Im nächsten Kapitel wechseln wir von der Bestätigung zur Theorienentwicklung. No-Go Theoreme sind wichtige methodologische Instrumente in der Theorienentwicklung. Sie beabsichtigen die Unmöglichkeit eines bestimmten Zieles darzustellen und treffen so explizite Aussagen über den Theorienraum. Ganze Forschungsprogramme wurden dadurch gestoppt. Aber sind No-go Theoreme wirklich so starke methodologische Werkzeuge? In Kapitel 4 beginne ich mit dem Beispiel eines No-Go Theorems, welches uns erlaubt eine abstrakte Definition von No-Go Theoremen zu geben. Wir zeigen, dass die Struktur von No-go Theoremen komplizierter ist als üblicherweise gedacht. Dies hat offensichtliche Auswirkungen auf ihre Interpretation. Die komplexere Argumentationsstruktur beinhaltet Elemente die garnicht oder nur schwierig zu rechtfertigen sind. Dies führt zu einer Interpretation, in der No-go Theoreme nie die Unmöglichkeit eines Zieles beweisen können, sondern am besten als methodologische Basis verstanden werden, von wo aus die Theorienentwicklung stattfinden kann. Das heisst, dass No-go Theoreme am besten als Go-Theoreme verstanden werden sollten.

Im letzten Kapitel haben wir uns mit einer anderen Möglichkeit auseinandergesetzt, um Theorien zu bestätigen. Schwer erreichbare Systeme wie schwarze Löcher werden heutzutage häufig mit Hilfe von Fluid-analogen Tischexperimenten simuliert. Die Frage ist, ob Belege, die wir in analogen Modellen erhalten als Belege für bestimmte Eigenschaften schwarzer Löcher betrachtet werden können? In diesem Fall könnte man das eine Modell durch Experimente am anderen Modell bestätigen. Dies ist normalerweise nicht der Fall. Der Grund hierfür ist der, dass alle Modelle auf bestimmten Annahmen basieren, die, wenn wahr, das Modell rechtfertigen. Die Annahmen in dem einem Model sind jedoch üblicherweise unabhängig von den Annahmen im anderen Modell. Damit können die Experimente im einen Modell nur die entsprechenden Annahmen bestätigen, nicht aber die des anderen Modells. Wir argumentieren jedoch, dass sogenannte model-externe, empirisch belegte Argumente (z.B. Universalitätsannahmen), welche die zwei Systeme verbinden, die Bestätigung des nicht zugänglichen Systems ermöglichen.

Chapter 1

Introduction

The history of the scientific method is a history of the relation between observation and theory. Ever since Aristotle it has been about careful observations of the world and the search for the principles that govern and explain it. While there is today no agreement about the 'right' scientific method, there has been significant progress in understanding the nature of scientific inquiry. But what if the observational side, or experimental side, of the relation would be missing? What if the consequences of the theories could not be tested? This seems to be the inevitable fate of fundamental physics and prompts the question of how, if at all, science can proceed from there.

Consider for instance the experimental development within particle physics, the search for the most fundamental particles and their interactions. The last 60 years have shown an incredible advance in accelerator technology. The very first accelerator at CERN in Geneva reached energies around 600 million electron volts (~ 1 GeV). The currently running Large Hadron Collider (LHC) at CERN reaches energies that go up to several trillion electron volts ($\sim 10^4$ GeV), using a 27 km accelerator ring and detectors the size of seven-story buildings. The LHC will make many precision measurements regarding the well confirmed Standard Model of particle physics, but also test several predictions of theories, which go beyond the Standard Model like supersymmetric extensions or extra-dimensional theories.

However, there are also many proposed theories at much higher scales. Grand unified theories are theories at 10^{16} GeV or theories of quantum gravity are Planck scale theories, i.e. at 10^{19} GeV. These are 12-15 orders of magnitude beyond what the LHC, the biggest experiment ever build, can reach. It is obvious that building bigger and bigger experiments will reach certain limitations, both financial and geographical. So testing these theories directly seems unrealistic in the foreseeable future and it needs to be replaced with indirect ways, like astrophysical experiments and precision measurements. These indirect methods of testing may provide hints of new physics, however, they are less controllable and have their own limitations as well (Bjorken, 2001; Mangano, 2007). So it seems an inevitable development of fundamental physics that there will be theories, which are perfectly fine scientific theories, that make predictions and can be falsified, but will not be testable due to practical limitations.

It is nevertheless the case that for several decades now scientists are strongly defending these untested theories. How can this strong commitment in spite of the lack of experiments be understood? Why do scientists trust their theories in the absence of empirical data? One reason that without doubt plays a crucial role in answering the above question is a sociological answer. These theories are defended by eminent theorists, much funding has been invested and there are many centres of e.g. String theory in the world. It is therefore not a surprise that scientists are not too willing to give up on their theories.

Another possible reason is that scientists are actually epistemically justified in trusting their theories even in the absence of empirical data.

One may argue, that there is a rationale that can be followed, which allows them to confirm theories either indirectly or even non-empirically, i.e. based on evidence which is not directly made more or less likely by the theory itself. In practice, scientists do informally use certain novel argumentative strategies to provide support for their theories and these seem to go beyond the purely sociological motivation to defend their theories. It is the aim of this dissertation to address some of these novel argumentative strategies. More explicitly, we will provide answers to the following questions:

- Is there an epistemically justified methodology of science in the absence of empirical data?
- If yes, are scientists on the basis of these methodologies, justified in trusting their theories?
- What are the capabilities and limitations of these novel methodologies?

By providing answers to these question, we will challenge the "orthodox" limitations on scientific methodology.

The main concern of this dissertation is scientific methodology in the absence of direct empirical data. More concretely, we are interested in the following situation. Consider a well-confirmed theory T from which a new theory or model T' is developed. The new theory T' makes novel predictions, however, these are for some reason empirically inaccessible. Within this setup I am interested in two issues:

Theory development: The process of developing theory T' from T.

Theory Assessment: Assessing theory T' if empirical data is rare or absent.

Theory development and theory assessment have been distinct aspects within the philosophical analysis of science. However, within the situation we are considering, i.e. where empirical evidence is lacking, it will become clear that in most cases they need to become closely intertwined.

Let us turn to the approach we would like to follow in this dissertation. The philosophical analysis of how we obtain and justify scientific knowledge has been concerned with the relation between empirical data and scientific theories that account for that data. Whether one was concerned with inferring inductively theories from data or confirming theories by testing their consequences, all these different approaches had in common that the theories' empirical predictions played the crucial role in assessing the scientific knowledge we have obtained through the theory. So how can we possibly assess theories, when we cannot directly test the empirical predictions of a theory?

Recently, Richard Dawid (2013) proposed in his book "String Theory and the Scientific Method" the possibility to confirm theories by assessing limitations on scientific underdetermination. The assessment works via observations, which are not made more or less likely by the theory one wishes to support. It is in this sense non-empirical. It was shown how, in principle, this kind of non-empirical evidence can confirm theories. The crucial element of these analyses is the concept of theory space, which is the concept we will be focusing on in our approach to evaluate theory development and theory assessment. All of the chapters in this dissertation will base their evaluation in one way or another on an evaluation of and exploration within theory space. This, as I aim to make plausible, provides a fruitful framework for philosophy of science more generally.

This dissertation contains four papers which address different aspects of theory development and theory assessment in the absence of empirical data.

In Chapter 2 we discuss the possibility to assess theories by assessing theory space. It is shown how non-empirical theory assessment can be understood as an extension of empirical theory assessment. While Dawid, does not propose to assess theory space explicitly, he provides arguments to indirectly assess theory space. These non-empirical assessments are, however, based on the assumption that scientists have explored theory space to a large extant. We discuss different problems for the possibility of theory space assessment and argue that if we want to assess theories by assessing theory space, scientific practice will have to change.

In Chapter 3 we will concentrate on one of the specific arguments provided by Dawid, namely the No Alternatives Argument. More concretely, it was claimed in (Dawid, Hartmann, and Sprenger, 2015), that the absence of alternative theories can confirm a theory. I will propose an amendment of the definition of the required evidence, in order to make it applicable to scientific theories. Within the changed setup, I address the possibility to confirm theories of quantum gravity on the basis of the No Alternatives Argument.

In Chapter 4 we discuss the consequences of the intricacies of theory space for the interpretation of no-go theorems. No-go theorems have been used in physics and elsewhere as important methodological tools in theory development. They make claims regarding what is and what is not possible in physics, having the effect of stopping whole research programmes. Impossibility claims are, however, explicit claims regarding theory space. I will argue that taking an adequate account of theory space into consideration will allow us to see how the effect of no-go theorems in the history of science is not warranted by the methodological implications they can have.

In the final Chapter 5, I consider the possibility for analogue simulation to be confirmatory. In analogue simulation the analogical relationship between the two systems is established via a syntactic isomorphism between the modelling frameworks used to describe each system. Such a relationships can be found, for example, between models of black holes and fluid mechanical analogue 'dumb holes' (Unruh, 1981). Significantly, within these pairs of systems, one half is typically experimentally inaccessible (e.g. black holes). We discuss the conditions under which evidence in the accessible system may count as evidence for the inaccessible system.

In the Conclusion I will present an overview of the results as well as offer further issues where the concept of theory space can be a crucial and also useful concept.

Chapter 2

Scientific Practice in Modern Fundamental Physics

2.1 Introduction

Consider the following three theories: cosmological inflation provides explanations, for instance, for the large-scale homogeneity and isotropicity of the universe, the flatness of the universe and the absence of magnetic monopoles. Supersymmetry extends the symmetry of the Standard model of particle physics by introducing for each particle we know a new supersymmetric partner. It solves many open problems, like the infamous hierarchy problem and provides a candidate for dark matter. String Theory is a proposed unified theory of all fundamental forces, i.e. gravity, the electromagnetic force, and the strong and weak nuclear forces. It is a theory most relevant at very high energy scales, the so-called Planck-scale, where one expects all fundamental forces to unify.

These three theories are exemplar theories in modern fundamental physics and they have two things in common. First, they all lack empirical support and second, they are, nevertheless, being defended by scientists for several decades despite the lack of empirical support. But why do scientists trust their theories in the absence of empirical data?

In a recent book Richard Dawid (2013) addresses this question and proposes an answer. The reason, he argues, is based on the idea that one can assess theories non-empirically by assessing the extent to which scientific underdetermination is limited. But how does non-empirical theory assessment fit into the predominantly empirical paradigm of scientific methodology and what kind of implications does this have for the practising scientist? In this paper we address (i) how non-empirical theory assessment can be understood as a natural continuation of empirical theory assessment and (ii) what the normative implications of non-empirical theory assessment are for scientific practice.

We start in Sect. 2.2 by giving a brief review of the empirical paradigm in scientific methodology and discuss how non-empirical theory assessment can be understood as complementing this. This is followed by Sect. 2.3, where we consider several problems for non-empirical theory assessment. These problems provide us with the normative implications for scientific practice, which we discuss in Sect. 2.4, before concluding in Sect. 2.5.

2.2 Scientific Methodology Without Experiments?

2.2.1 Scientific Methodology: The Empirical Paradigm

Scientific theories are supposed to provide us with an empirically adequate account of the world. It is therefore not surprising that when we want to assess them, we confront them with the world.[1] The history of

[1] I use "assess" when I do not want to consider a specific kind of assessment, i.e. I leave it open whether the assessment is via Popperian corroboration, H-D confirmation, Bayesian confirmation etc.

the scientific method has, therefore, been a history of finding the right methodology to relate empirical data and theories.[2]

Very broadly construed one can understand the history of the scientific method in two parts. A pre-20th century inductivist tradition developed by the likes of Aristotle, Bacon, Newton or Mill and a post-20th century shift in methodology towards an assessment via the consequences of the theories (Fig.2.1). The aim of the inductivist tradition was to identify reliable methods to infer inductively from observations to generalisations.

The situation changed post-20th century. The revolutions that occurred in physics, like the development of quantum mechanics and the special and general theory of relativity, showed the strong fallibility involved in the inductive method. If even Newtonian mechanics could be wrong then it may be the case that the inductive method is not as reliable as we thought. So rather than concentrating on how to develop a theory out of empirical data, one should start with the theory and consider its consequences. It did not matter, whether the inductive method was used in the development of the theory or whether the theory appeared in someone's dream. Popper, for instance, proposed an account of corroboration and falsification, followed by theories of confirmation culminating in modern Bayesian confirmation theory.

While these methodologies are quite different, they all aim to provide a normative component. That is, they do not always aim to provide the framework to rationally reconstruct how scientists obtain scientific knowledge but also provide the rules and guidelines of how they should go about obtaining it. Let us consider some examples. As mentioned, the pre-20th century philosophers and scientists aimed to identify different

[2] See (Nola and Sankey, 2014) for a textbook introduction or (Andersen and Hepburn, 2015) for an overview article on the scientific method.

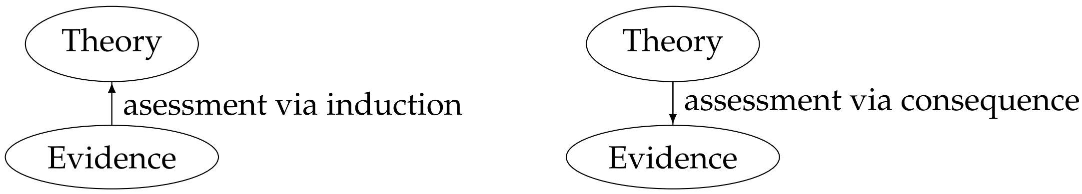

FIGURE 2.1: Left: Simplified Scheme of the direction of inference in assessing theories pre-20th century. In the inductivist tradition one developed reliable methods of inference from observations to generalisations. Right: In post-20th century scientific methodology the focus shifted towards an assessment via the consequences of the theory.

methods of inferring from observations to their generalisations. To give an example, Newton formulated certain rules that scientists, i.e. natural philosophers at that time, should follow. His Rule III, for instance, says:

> Those qualities of bodies that cannot be intended and remitted [i.e. qualities that cannot be increased or diminished] and that belong to all bodies on which experiments can be made, *should* be taken as qualities of all bodies universally. [Quoted in (Harper, 2011, p.272), my emphasis.]

For John Stuart Mill, induction amounted to finding regularities in the observations. He tried to identify methods by which to identify causes.[3] He proposes in his "Systems of Logic", for example, his famous method of difference:

> Second Canon: If an instance in which the phenomenon under investigation occurs, and an instance in which it does not occur, have every circumstance save one in common, that one occurring only in the former; the circumstance in which alone the two instances differ, is the effect, or cause, or a necessary part of the cause, of the phenomenon. (Mill, 1843, p.483)

[3] See for instance (Wilson, 2016, Sect. 5).

Popper within the post-20th century tradition claims "[e]very genuine test of a theory is an attempt to falsify it, or to refute it." and that "[c]onfirming evidence should not count except when it is the result of a genuine test of the theory" (Popper, 1989, p.36). So rather than showing that your theory makes correct predictions one should focus scientific practice on trying to refute the theory.

The details or the viability of these normative claims do not matter at this point. However, all have in common that if taken seriously they have direct implications for scientific practice. That is, they are not solely (or at all) descriptive accounts of how science generates new knowledge but normative accounts of how scientific practice should proceed in order to reliably produce scientific knowledge.

2.2.2 Scientific Methodology 2.0: Towards A Non-Empirical Paradigm in the 21st Century

The scientific revolutions in early 20th century physics gave rise to a re-evaluation of scientific methodology. A change, one may safely say, was not accompanied by a change in actual scientific practice. What has changed that would make it necessary to reconsider scientific methodology in the 21st century? The scientific theories in fundamental physics are not more detached from our intuition as it was the case in the early 20th century when quantum mechanics and the theories of relativity were developed. So the problem does not lie necessarily with the theories themselves but with the situation regarding empirical data. For most theories beyond the standard model of particle physics and theories of quantum gravity, empirical data is either rare or completely absent. However, String theory, supersymmetry or cosmic inflation are all theories that are defended for decades now in the absence of empirical data to support them. Why? Non of the methodologies developed in the previous two millennia seem to apply .

One possible route, is to use external criteria, which are not based on empirical evidence (Fig.2.2). An example is the assessment of theories based on aesthetic criteria. One proponent for such a view is Paul Dirac and his use of mathematical beauty to assess theories.[4] When Dirac claims "[t]he research worker, in his efforts to express the fundamental laws of Nature in mathematical form, should strive mainly for mathematical beauty" (Dirac, 1940, p.123), he provides us with a methodology of theory development and assessment. He proposes the position that

> [i]t is more important to have beauty in one's equations than to have them fit experiment. [...]. It seems that if one is working from the point of view of getting beauty in one's equations, and if one has really a sound insight, one is on a sure line of progress. (Dirac, 1963, p. 47)

This has, of course normative consequences. On the one hand, in comparison with other aesthetic criteria, where he claims that "[scientists] should [...] take simplicity into consideration in a subordinate way to beauty"(ibid.). On the other hand, in comparison with empirical data, where he claims with respect to the general theory of relativity that

> [t]he foundations of the theory are, I believe, stronger than what one could get simply from the support of experimental evidence. The real foundations come from the great beauty of the theory. (Dirac, 1980, p.10)

[4] See (McAllister, 1990) for a philosophical perspective on Dirac's aesthetic criteria.

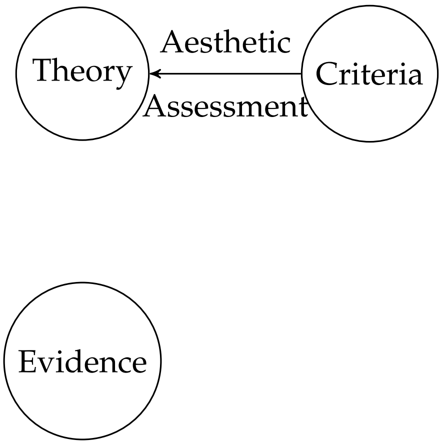

FIGURE 2.2: Many theories in fundamental physics, especially theories of quantum gravity, currently find themselves in a situation where one can neither infer the theory from empirical data nor test the theory via its consequences. That is non of the previously developed methodologies are applicable. Some propose a methodology based on aesthetic criteria.

However, there are not many supporters of this methodology.[5] The main reason for this is the lack of an objective account of beauty.[6] There is not a unique explication and as such it is not clear how it should correlate with truth. Moreover, it is unclear how one can obtain evidence for the claim that beauty and truth correlate in the first place.

We, therefore, want to turn to a more conservative and more defendable position, namely Richard Dawid's account of non-empirical theory assessment. He proposes that it is possible to assess theories even in the absence of empirical evidence, by what he calls non-empirical evidence. That is, one can confirm theories even if non of the novel empirical consequences of the theory can be tested.

The assumption that one can assess theories non-empirically seems at

[5]McAllister (1999) defends the most sophisticated account of the relation between truth and beauty within his aesthetic induction account. See also (Kuipers, 2002).

[6]A point Dirac accepts to some extent, when he denies the possibility to define beauty: "[Mathematical beauty] is a quality which cannot be defined, any more than beauty in art can be defined, but which people who study mathematics usually have no difficulty in appreciating" (Dirac, 1940, p.123).

least counter-intuitive, especially since we aim to find empirically adequate theories. So how then can Dawid's *non-empirical* theory assessment tell us anything about the empirical adequacy of a theory? Let us take a step back and consider one possible line of reasoning. Consider Newton's law of gravity

$$F = G\frac{m_1 m_2}{r^2}, \tag{2.1}$$

where F is the force acting between the two masses m_1 and m_2, G is the gravitational constant and r the distance between the two masses. Once we fix m_1 and m_2 and determine the distance between them, we can calculate the gravitational force acting on them. Using a dynamometer we can now measure the force and see whether it agrees with the prediction. In case it does, the hypothetic-deductive as well as Bayesian confirmation theory will tell us that we have confirmed (2.1).[7] There seems to be no doubt that we have provided empirical support for Newton's law of gravity.

However, with any confirmation comes a certain expectation. The expectation corresponds to us trusting the predictions of the theory. But we may trust some of the predictions more than the other. This difference is determined by what we take our background knowledge to be. To be more concrete, while I may have confirmed Newton's law of gravity and so have increased my subjective degree of belief with respect to it, I do not assign equal probability to all of its possible consequences. Imagine, for example, testing Newton's law of gravity for $m_1 = 10$ kg and for $m_1 = 12$ kg, keeping everything else from above fixed. Let us assume, that in both cases we confirm (2.1). These two examples of empirical evidence together with our background knowledge provide us with the

[7] From now, when we talk of assessment we mean confirmation in the Bayesian sense and mean with "confirm (2.1)", the confirmation of the proposition stating that (2.1) is empirically adequate.

expectation that in case we would test the theory with $m_1 = 11$ kg we would still get the right prediction. It seems we have good reason to make that inductive leap and trust the theory for this untested value as well. But would we have the same expectation for $m_1 = 10^{55}$ kg? Would we equally well be justified to make the inductive leap and trust (2.1) for that value, and if not, why?

Without any background knowledge we do not have any reason to have different expectations for the empirical adequacy of Newton's law of gravity for small values of m_1 compared to large values of m_1. However, we do have additional background knowledge, so that we may believe that we can trust (2.1) for small values of m_1, while not for values much higher. This background knowledge, and this is the crucial point, can equivalently be understood as constraints in theory development. Let us illustrate this. Why do I trust Newton's law of gravity for the unconfirmed case of $m_1 = 11$ kg? One plausible reason would be that any alternative theory, which I could develop would need to agree with (2.1) for the values $m_1 = 10$ kg and for $m_1 = 12$ kg while deviating from it in between. This seems incredibly unintuitive and would violate what Dawid (2013, Sec.3.2) calls scientificality conditions. These are certain ampliative rules of theory development. Following (Laudan, 1996) examples would be the exclusion of ad-hoc explanations for individual events, simplicity assumptions, etc. That is, the scientificality conditions, if we accept them, put constraints on possible theories we could build, which would deviate from Newton's law of gravity at $m_1 = 11$ kg. In this case, they would make alternative theories, which would deviate from Newton's law of gravity at $m_1 = 11$ kg very unlikely. That is why we trust Newton's law for that value and that is also why we build bridges trusting the laws will hold also for those untested values.

So why is our expectation with regard to $m_1 = 10^{55}$ kg different? Well, first, the scientificality conditions do not equally well apply to the formulations of alternative theories for those high masses. Deviations would seem much less ad hoc. That is, the scientificality conditions' constraint on alternative theories at higher values seems less stringent. Second, and more importantly, we have an alternative theory, which makes different predictions for those high values, namely Einstein's theory of General Relativity. The two values from above, which were used to confirm Newton's law of gravity, similarly confirm General Relativity. That is, all the evidence that supports one theory also supports the other.[8] We do not need to commit to either of the theories, for the values they agree on. That they agree for certain values, actually should increase our trust in the predictions for those values. However, that the predictions disagree for large mass values will have an effect on our expectations regarding the predictions of these theories for those values. This already shows how an assessment of theory space, i.e. an assessment of the number of alternatives, provides a strategy to assess how much of a theory is being confirmed by any empirical evidence.

But this is not yet the end of the story. We could actually go on and consider whether there are further reasons to trust one theory over the other for those high values. One option would be to consider further theoretical constraints. One may want to argue that the requirement of general covariance will constrain theory space further and leave us only with General Relativity. Another option would be to argue that General

[8] There is a possible issue with respect to the question of how much of Newton's law of gravity was assumed in the development of General Relativity. Let us assume it was not used in the development for the purpose of this argument, as we only provide a plausibility argument here and not a claim with respect to General Relativity vs. Newton's law of gravity.

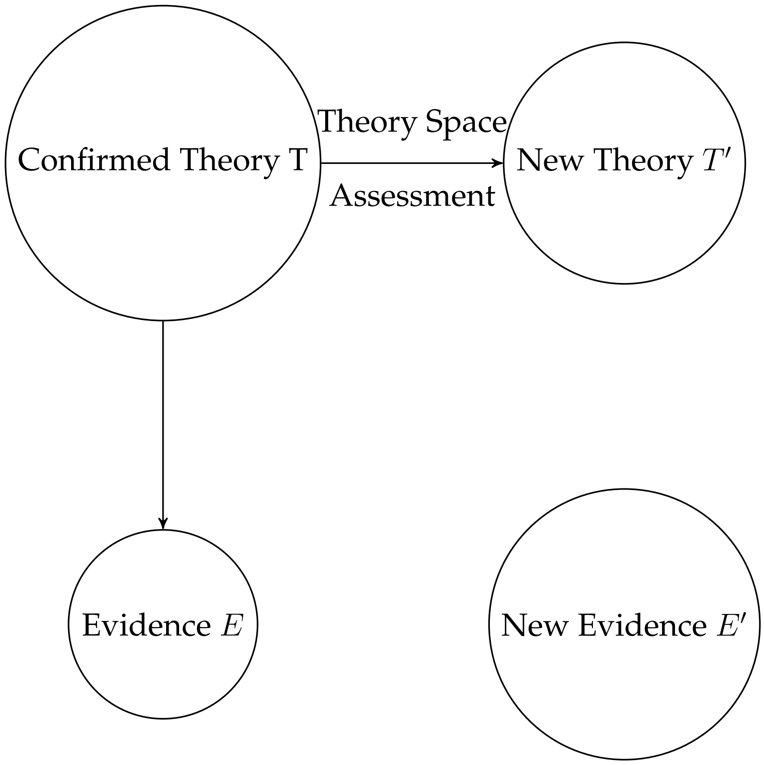

FIGURE 2.3: The non-empirical strategy: Assessment of the new theory by an assessment of theory space.

Relativity is related in unexpected ways to other theories, which are confirmed etc. If this is the case, it seems reasonable to assume that this will have an impact on our expectations. None of these have to be the case and much of the viability of this approach will depend on whether these additional constraints on theory space can be epistemically justified. However, they do provide a general strategy of non-empirical theory assessment. But this is, of course, only the case, if we can actually probe theory space to see how constrained it is, or to use Dawid's term, how strong the limitations on scientific underdetermination[9] are.

Theory space is a very difficult concept to make precise, so how can we possibly have access to it, in order to analyse how constrained it is. Dawid discusses three arguments that, in case one can establish them,

[9]The notion of underdetermination used by Dawid (2013, Sect. 2.2) refers to the underdetermination of empirically inequivalent theories with respect to existing evidence, which by further evidence can be confirmed differently. The notion has strong similarities with transient underdetermination used in (Sklar, 1975) and (Stanford, 2006, p.17)

provide non-empirical evidence for the limitation on scientific underdetermination and thereby for the theory (Dawid, 2013, Sect. 3.1). What is non-empirical about this kind of evidence? Well, neither the empirical data E supporting the confirmed theory T nor the possible new data E', which would possibly confirm T' directly, is used to test the theory. The non-empirical evidence used is solely for the purpose of gaining insight about how constrained theory space is and is not made more or less likely by T' itself.

One of the specific arguments provided by Dawid is the No Alternatives Argument (NAA). Consider the following observation: "the scientific community, despite considerable effort, has not yet found an alternative to theory T fulfilling certain constraints" (Dawid, Hartmann, and Sprenger, 2015, p.217). One can understand this observation as providing non-empirical evidence for there not being any alternatives or at least a very limited number of them. This by itself is a weak argument and bears many problems (See Chapter 3). It is only the conjunction of all three arguments that provide relatively strong evidence for limitations on scientific underdetermination. The details of the other two arguments, however, do not concern us for the purpose of this paper as we are more concerned with the general problem of exploring theory space.

But before we turn our attention to the intricacies of theory space assessment, let us consider a possible worry one may have with the assumption that one can assess theory space at all. Kyle Stanford argues explicitly against the possibility to conceive of possible alternatives. He starts with the following claim:

> I suggest that the historical record offers plainspoken inductive testimony to the fact that we have repeatedly occupied a predicament of recurrent, transient underdetermination across a wide and heterogeneous variety of scientific fields and domains of inquiry [...]. (Stanford, 2006, p.19)

He goes on to use this historical evidence in his "new induction" to argue against a realist account of theories. While the metaphysical component of the argument does not concern us at this point, a worry remains. This becomes clear in the following claim that

> [...] in the past we have repeatedly failed to exhaust the space of fundamentally distinct theoretical possibilities that were well confirmed by the existing evidence, and that we have every reason to believe that we are probably also failing to exhaust the space of such alternatives that are well confirmed by the evidence we have at present. (Stanford, 2016)

So, he argues, (i) that there is historical evidence that we should expect their to be alternative theories and (ii) that these are usually unconceived by the scientists at any given time. Both (i) and (ii) seem to pose problems for non-empirical theory assessment.

Let us start with claim (i). Let us assume, with Stanford, that we have historical evidence that we should expect there to be alternative theories. That is, whenever we have a theory, there will be alternative theories able to satisfy the same constraints. If this is the case, we simply will not be able to obtain non-empirical evidence pointing to limitations on scientific underdetermination. So if this is true, there simply will not be many cases, where we will be able to apply the non-empirical methodology. This is not a problem of principle for non-empirical theory assessment but only one of limited applicability. Let us turn to (ii). If the alternative theories are unconceived this is definitely worrisome, because it shows that it is not an obvious task to assess the number of alternatives. But it, of course, does not entail that the alternative theories are unconceivable. There are at least two reasons for that. First, the historical evidence that there are unconceived alternatives is based on the evidence that later on the alternatives were found. And second, why should scientists have looked for alternatives? If a theory already exists that satisfies all the constraints, why would a scientist look for alternatives? It does not seem to be the case that the search for alternative theories competing with the existing successful theory is high up on the scientist's priority list. Only when e.g. anomalies appear, do scientists have reason to look for alternatives. And when they actually do, they do come up with distinct theoretical possibilities that were well confirmed by the existing evidence. So, to turn the argument around, one may argue that we have good reason to believe that if scientist search for alternatives they will find them. It is only that they usually do not look for alternatives until it seems necessary.

It is this last point, which is crucial for the rest of the paper. Scientists do not actively search for alternative theories, but this is crucial for non-empirical theory assessment. When Dawid states in the case of the non-empirical evidence of the NAA that scientists do not find alternative theories "despite considerable effort", it is crucial that this "effort" is actually real. That is in order to assess theories non-empirically, scientists will have to change their focus in research. This provides a first hint at the normative implications of non-empirical theory assessment for scientific practice. In the rest of the paper we discuss several problems scientist can encounter in assessing theory space.

2.3 Problems of Theory Space Assessment

In the following we will address several problems that arise in determining a theories status with regard to its position in theory space. The approach is to consider specific cases from the history of physics where scientists have mistakenly constrained theory space. This will allow us to identify what the elements are that constrain theory space and what the possible pitfalls are in using them prematurely.

2.3.1 The Theoretical Problem or Quo Vadis, Theory Space?

Georgi and Glashow proposed in 1974 a grand unified theory based on the mathematical group $SU(5)$. They provide a unification of all fundamental interactions of particle physics. They start the paper by claiming

> "We present a series of hypotheses and speculations leading inescapably to the conclusion that $SU(5)$ is the gauge group of the world." (Georgi and Glashow, 1974, p.438)

They then go on to develop the theory and end with the claim:

> "From simple beginnings we have constructed the unique simple theory." (Georgi and Glashow, 1974, p.440)

So here we have an example, where scientists make explicit statements with regard to theory space. They say they have provided a theory, which "inescapably" contains $SU(5)$ as the "gauge group of the *world*"[10] and it is a "unique" theory. Whether they meant it as strongly as it is suggested by these quotes is not relevant for our purposes. We are only interested

[10] My emphasis.

in reconstructing the necessary constraints that would lead to the conclusion they draw.

Let us start with the gauge group of the standard model of particle physics from which Georgi and Glashow develop their $SU(5)$ theory. The gauge group of the standard model is $SU(3)_C \times SU(2)_W \times U(1)_Y$. The particle content in their respective representations of the gauge group is listed in Table 2.1. What is important is that any future theory should

Names		$SU(3)_C \times SU(2)_W \times U(1)_Y$
Quarks	$Q = \begin{pmatrix} u \\ d \end{pmatrix}$	$(\mathbf{3}, \mathbf{2})_{+1/3}$
	u^c	$(\bar{\mathbf{3}}, \mathbf{1})_{-4/3}$
	d^c	$(\bar{\mathbf{3}}, \mathbf{1})_{+2/3}$
Leptons	$L = \begin{pmatrix} \nu_e \\ e \end{pmatrix}$	$(\mathbf{1}, \mathbf{2})_{-1}$
	e^c	$(\mathbf{1}, \mathbf{1})_{+2}$
	$[\ \nu_e^c$	$(\mathbf{1}, \mathbf{1})_0]$
Gauge Bosons	G	$(\mathbf{8}, \mathbf{1})_0$
	$A^{\pm}, A^3$	$(\mathbf{1}, \mathbf{3})_0$
	B	$(\mathbf{1}, \mathbf{1})_0$
Higgs	ϕ	$(\mathbf{1}, \mathbf{2})_{+1}$

TABLE 2.1: The Standard model gauge group and its particle content (first generation only) in their respective representations. Any future theory of particle physics should contain these. See e.g. (Griffiths, 2008).

accommodate the particle content of the standard model and contain the gauge group as a subgroup. This is crucial to guarantee the empirical adequacy of any future theory with respect to the evidence that confirmed the Standard model already.

The standard model gauge group is a Lie group. Luckily, simple Lie groups have been completely classified. That is we have a fixed set of possible groups to choose from. Let us consider several of the constraints on any future gauge group G (Georgi, 1999, p.231-234):

(1) Group G must be rank ≥ 4: this is needed to contain the four commuting generators of the standard model gauge group.

(2) Group G must have complex representations: some of the particles, like u^c, are in the complex representation of $SU(3)_C$ and so need to be accommodated by any future group.

(3) Group G should be a simple[11] group: this ensures that the gauge couplings are related.

(1)-(3) will put strong constraints on the set of possible groups that come into consideration. For instance, if we restrict the groups to simple groups we rule out for instance groups like $SO(7)$ (rank 3) or F_4 and $SO(13)$ (both no complex representation) to mention a few. However, groups like $SU(5)$, $SU(6)$, $SO(10)$ and many others remain. So it seems not in any way to lead us "inescapably" to $SU(5)$ as the gauge group "of the world". The only way $SU(5)$ could be considered the unique group, is by adding an additional constraint:

(4) Group G should be the simplest group satisfying (1)-(3).

So given constraints (1)-(4) we may say that theory space allows for only one possible group, namely $SU(5)$.

Let us start by evaluating the different constraints. It is obvious that the different constraints differ in strength of justification one can give for them. It seems restriction (1) and (2) are reasonably supported theoretical constraints on G, as they represent minimal requirements to account for the successes of the standard model gauge group. That is they are empirically supported. On the other hand, (3) seems not to be necessary in a similar sense. Having a simple group has the nice feature that the three gauge couplings of the standard model are then related to each other after the symmetry is spontaneously broken. There is however, nothing empirically requiring this to be the case. Slansky, for example, even says with

[11] A group is simple, if it does not contain any invariant subgroup.

regard to simple groups that the "restriction is physically quite arbitrary" (Slansky, 1981, p.14). A couple of months after Georgi and Glashow, Pati and Salam (1974) actually provided a unification based on a group, which was not simple. The last constraint we considered, that we pick the simplest out of the simple groups, lacks any reasonable justification other than possibly aesthetic ones. So we see, how one may have been led to think one has found a "unique" theory but this assessment will only be as strong as the constraints that led to the uniqueness. At least in this case not all constraints are well justified.

There is a further problem. While we saw why we may need to worry about wrong constraints, we may also need to worry about the very problem the theory aims to solve. Consider for a moment that all the theoretical constraints from before are empirically well confirmed (imagine this is possible) and based on these constraints there are no alternatives. We would be lead to believe that this theory is empirically adequate, as there simply are not any other theories. However, it can now simply be the case that the problem we are considering is simply not a genuine problem. Unlike empirical problems, conceptual problems like unification are on shaky grounds. In the above example one considers the problem in need of a solution to be the unification of all the standard model interactions. But how do we know this is really a problem?[12]. The danger is that we find a theory, which has no alternatives to a problem that is not really a problem. In this case we would have developed a theory towards the wrong "direction" of theory space.

Physicists have the impulse to develop theories without rigorously trying to justify every step of the way all the assumptions involved in

[12]We will get back to this problem within the context of the No Alternatives Argument in Chapter 3.

theory construction. This is, of course, fine if one can then do experiments to test the theory. Georgi and Glashow's $SU(5)$ theory predicted, for instance, that the proton would decay. However, there has not been any observation of proton decay and the bound put on the lifetime of the proton disagrees with the predictions of $SU(5)$. If, however, experiments are not possible, and one is left with non-empirical theory assessment as the only means to test the theory, one needs to (i) carefully assess all the constraints used in the development as well as (ii) provide good reason that the problem is genuine. These are necessary ingredients for the reliability of non-empirical theory assessment.

2.3.2 The Structure Problem or Where is the Constraining Taking Place?

Consider some physical assumption. While one may phrase that physical assumption colloquially in words, one usually has a more precise formal representation of it in mind. When I, for example, talk of probabilities and certain features they should satisfy, I have more formally the mathematical structure of Kolmogorovian probabilities in mind, which satisfy certain specific mathematical axioms. That is I link the physical concepts with specific mathematical structures, which represent them. In the previous example we considered symmetries of some internal space, which are represented in terms of the mathematical structure of Lie groups. The conceptual problem of unification, together with the other constraints, like the particle content and their representations, then translated into the attempt to unify these different Lie groups, into one bigger Lie group. The assessment of theory space was explicitly determined by the classification of all simple Lie groups. That is we could just rule out certain points in theory space by ruling out Lie groups that do not satisfy the constraints. But is this the right place for the constraining to take place?

To illustrate the problem that arises with this let us consider an example. The strongest kind of assessment of theory space are impossibility results. One proves that, given certain constraints, there are no theories in theory space that are able to satisfy them. In an example we discuss in more detail in Chapter 4, scientists aimed to find a unification of internal and external symmetries. They did not only fail but even claimed it is impossible. Later, and for independent reasons, scientists were actually able to unify internal and external symmetries. Not by changing certain physical constraints but by changing the mathematical structures used. The mathematical structure used in the impossibility results were Lie algebras. It is instructive to consider it in more detail. Let us quickly remind ourselves of some definitions[13].

> A **Lie Algebra** consists of a vector space L over a field ($\mathbb{R}$ or $\mathbb{C}$) with a composition rule, denoted by $\circ$, defined as follows:
>
> $$\circ : L \times L \to L$$
>
> if $v_1, v_2, v_3 \in L$, then the following properties define the Lie algebra:
>
> 1. $v_1 \circ v_2 \in L$ (closure)
> 2. $v_1 \circ (v_2 + v_3) = v_1 \circ v_2 + v_1 \circ v_3$ (linearity)
> 3. $v_1 \circ v_2 = -v_2 \circ v_1$ (antisymmetry)
> 4. $v_1 \circ (v_2 \circ v_3) + v_3 \circ (v_1 \circ v_2) + v_2 \circ (v_3 \circ v_1) = 0$ (Jacobi-identity)

[13] See (Corwin, Ne'eman, and Sternberg, 1975) for a review article and (Kalka and Soff, 1997) and (Mueller-Kirsten and Wiedemann, 1987) for elementary discussions.

Both the internal and external symmetries were represented as Lie algebras. The aim, similar to the previous example, was to find a Lie algebra that would bring them together. However, mathematicians already in the mid-1950s found a more general structure. The idea is that one can circumvent the pure commutator structure apparent in the anti-symmetry claim by defining the vector space as the direct sum of two vector spaces with different composition properties. This is, for example, achieved in what one calls $\mathbb{Z}_2$-graded Lie algebras:

> A $\mathbb{Z}_2$**-Graded Lie Algebra** consists of a vector space L which is a direct sum of two subspaces L_0 and L_1, i.e. $L = L_0 \oplus L_1$ with a composition rule, denoted by $\circ$, defined as follows:
>
> $$\circ : L \times L \to L$$
>
> satisfying the following properties:
>
> 1. $v_i \circ v_j \in L_{i+jmod2}$ (Grading)
> 2. $v_i \circ v_j = -(-1)^{ij} v_j \circ v_i$ (supersymmetrisation)
> 3. $v_i \circ (v_j \circ v_l)(-1)^{il} + v_l \circ (v_i \circ v_j)(-1)^{lj} + v_j \circ (v_l \circ v_i)(-1)^{ji} = 0$ (Gen. Jacobi-identity)
>
> with $v_i \in L_i$ (i=0,1).

This new structure has several interesting features. First, the grading gives rise to the feature that the composition of two L_0 elements gives an L_0 element, the composition of an L_0 element with an L_1 element gives an L_1 element and finally the composition of two L_1 elements gives an L_0 element. So L_1 by itself is not even an algebra, since it is not closed, while L_0 by itself is. Second, the so-called supersymmetrisation leads to a commutator composition for all cases but when two elements are taken from L_1. In those cases the composition is given by an anti-commutator. From this it follows that Lie algebras are special cases of graded Lie algebras in the case where L_1 is empty. This generalisation of the mathematical structure simply allows us to do more. In the above example, the non-trivial unification of internal and external symmetries is now possible with this new mathematical structure. This is problematic, because it seems that any assessment of theory space will depend on the choice of the mathematical structure within which the different theoretical and empirical constraints are assessed. This is what we call the *Structure Problem*. The choice of mathematical structure, more so than the theoretical constraints, is highly problematic with crucial implications for the interpretation of no-go theorems, as we will see in Chapter 4.

2.3.3 The Data Problem or Where to Start in Theory Space?

A final and maybe less obvious problem is, what I would like to call, the *data problem*. There seems to be nothing more uncontroversial then the constraints that come from empirical data. But what one may consider as empirical data is far from obvious. Take one apparently obvious example, the dimensionality of space. One may say it suffices to open the eyes to see that as an empirical fact space is three-dimensional. The three-dimensionality of space is a strong, if not the strongest, empirical constraint on physical theories. In formulating a theory it seems out of question to simply start in three space dimensions. Even with the advent of the general theory of relativity, where space itself became a dynamic entity, $D = 3$ was not under discussion. However, starting in theory development with the strong empirical constraint of $D = 3$ may put too strong of a constraint on theory space. Let us consider this example a little bit further, by considering examples from physics where extra-dimensions were introduced.

Unsurprisingly, they did not consider higher dimensional theories for the sole purpose of having a higher dimensional theory. They had different motivations. For instance, Nordström (1914) and Kaluza (1921) were hoping to unify gravity with electromagnetism and realised that a theoretical option would be to introduce an additional space dimension.[14] More recently Arkani-Hamed, Dimopoulos, and Dvali (1998) proposed a solution to the hierarchy problem, the unexplained difference in strength between the electroweak force and the gravitational force, by introducing curled up extra-dimensions. The idea is that the gravitational constant is a dimension-dependent quantity and in case we have more dimensions we can change the gravitational strength and thereby bring it closer to the electroweak scale. In case the radius of the curled up dimension is chosen to be small, we might not have been able to observe it by any experiment. At the time of the proposal, the extra-dimensions could even have been of millimetre size.

At this point one may think that, well, sure, one may not be able to rule out curled up extra-dimensions by observation, but we can at least be sure that there are no additional extended non-compactified extra-dimensions. Randall and Sundrum (1999), just like Arkani-Hamed et al. tried to solve the hierarchy problem and showed that one does not need to require the extra-dimensions to be curled up. As long as the electroweak interactions are constrained to three-dimensional space, it is possible for the additional dimensions to be extended. The gravitational

[14] Since Einstein had not yet developed his theory of general relativity in 1914, Nordström used his own empirically inferior scalar theory of gravity.

interactions, unlike the electroweak interactions, extends into these additional dimensions and that explains why it is weaker. Again, at the time of the proposal, there was no empirical evidence to rule these higher dimensional spaces out. So although one may have thought that the dimensionality of space is an empirical fact in no need for any further justification, it turned out to be much more complicated.

The three-dimensionality of space, a seemingly obvious empirical fact about nature, turned out to be an incredibly flexible element in theory development. When we observe the world around us we see three space dimensions. But as our vision is restricted to only electromagnetic interactions with the world and as the resolution of the eye is approximately one arc minute[15] there is plenty of leeway for theory development. Further experiments will, of course, put further limits on this leeway but will never rule them out irrefutably. It is important to realise that this is not a unique example but is a general feature of empirical data that is used in theory development. The empirical data used by scientists in theory development never consists of protocol statements in the Carnapian sense but extrapolations thereof. Whether we consider the tri-dimensionality of space or the homogeneity and isotropicity of the universe, none of these are irrefutable empirical statements.

2.4 The Normative Impact on Scientific Practice

In science we find both reliable and unreliable methods of theory development and assessment. When scientists and philosophers of science develop scientific methodologies they want to identify the reliable methods.

[15] One arcminute corresponds to a resolution of about a millimetre at a distance of 30 cm.

Once you consider your methodology reliable, you are imposing normative rules on scientific practice to guarantee that the scientists follow the reliable method as closely as possible.

Non-empirical theory assessment is based on an assessment of theory space. We have considered several examples from physics to illustrate different ways one can mistakenly constrain theory space. In these cases, one considered theory space to be more constrained then what was warranted by the available evidence. Recognising a problem is the first step in addressing it. If we want to assess theories based on non-empirical theory assessment we need to address these problems as they threaten the reliability of the non-empirical evidence we obtain. The success of non-empirical theory assessment will therefore strongly depend on how well we can address these problems.

We saw that by developing a theory we need to make certain assumptions to get started. Based on empirical, theoretical and mathematical assumptions we develop a theory able to solve some problem. Whether these are well justified or not, does not matter as long as there is empirical evidence that can be used to test the theory. If, however, there is no empirical evidence, it is crucial to consider the legitimacy of each constraint and assumption. An assessment of theory space to a large extent depends on the legitimacy of the assumptions involved in theory development and the constraints they impose. The theoretical, structure and data problem of non-empirical theory assessment suggest that it will be very difficult to assess theory space. The question of how to address these problems needs to be addressed case by case. It will require a careful analysis of how the constraints constraint theory space in each case and the development of possible strategies of inductive justification of the constraints involved.[16]

At this point we cannot provide detailed claims regarding how to best address the problems of non-empirical theory assessment in each case. This would require a more detailed analysis as we have provided above, however, we can recognise some general consequences of non-empirical theory assessment for scientific practice. In order to assess theories non-empirically a conscious shift of focus in scientific practice is needed. The most obvious change of perspective in scientific practice will be the focus on what is usually called the context of discovery. Rather than coming up with new theories and then considering their empirical consequences, the focus should be on justifying the very assumptions that led to the theory in the first place. Let us consider each of the problems in turn.

It is important to consider the scientific problems one wishes to solve carefully. If, for instance, I have some fine-tuned element in my theory, I may want to try to solve it, but one should also recognise the possibility of contingent unexplainable elements of theories. Or at least the possibility that the theory at the next scale does not yet provide the solution to the fine-tuning problem. This, of course, makes it necessary to analyse the very question of what constitutes a genuine scientific problem. A highly non-trivial problem. On the other hand one has to be aware of imposing strong theoretical constraints, even when they are well-confirmed. Consider for instance the theoretical principle of CPT-symmetry or Lorentz-invariance. Lorentz-invariance has been a successful ingredient of many well confirmed theories in physics, from classical electrodynamics via special relativity up to the standard model of particle physics. This may provide good reasons to consider it a meta-principle

[16]Some of these issues will be discussed further in Chapters 3 and 4

that needs to be required of all theories. However, this again may prematurely constrain theory space. Therefore, an approach has been to test violations of the Lorentz symmetry, by e.g. considering preferred frame effects, by extending the Standard Model of particle physics with Lorentz-symmetry breaking operators or by what is called very special relativity, where it is shown that only a subgroup of the Lorentz group is needed to account for all the standard predictions (Mattingly, 2005; Liberati, 2013). These are explicit methods to address the problem of constraining theory space based on unwarranted theoretical constraints. In this case there has been an extensive set of experiments testing various possible violations of Lorentz invariance within the different proposed frameworks (Russell and Kostelecky, 2009). So if the theory one is developing is at an energy scale where the empirical evidence does not show a violation of Lorentz invariance, requiring the Lorentz symmetry, and thereby constraining theory space, is empirically justified. If, on the other hand, I am developing a theory several orders of magnitude beyond the empirical bounds, theory space is not yet constrained by the data, and one may prematurely constrain theory space by requiring it. These are possible effective methods that are required to test the viability of theoretical constraints. Scientists, in the absence of empirical data to support their theories, should focus on these scientific practices to epistemically justify their guiding principles.

Let us turn to the structure problem. This problem seems less accessible. The problem is that scientists in practice do not recognise it as a constraint. When trying to solve some problem, one uses, understandably, the mathematical structures one always used. They were successful in the past, why should they not be in the future? Only if we recognise that we cannot solve a problem by using that structure, one may look for an alternative structure that allows one to solve the problem. If one finds a structure one stops looking, as there is no need to continue the search as one has solved the problem. Once one has solved the problem, with a specific mathematical structure their is no incentive in continuing the search for other mathematical structures that may also do the job. In this case one may be lead to think, based on the well supported theoretical constraints etc. that theory space is constraint. However, now there is an incentive. In case I do not have empirical data to test my theory, I may want to rely on theory space assessment. So if I would like to argue in favour of my theory I need to actively pursue the possibility of alternative mathematical structures. As we mentioned, the mathematical structure of graded Lie algebras was developed in the mid-1950s. If, O'Raifeartaigh or Coleman and Mandula would have actively searched for alternative mathematical structures, even within the existing mathematical literature, they would have recognised their unwarranted constraint on theory space.

The data problem can be addressed similarly to how the problem of theoretical constraints can be addressed. Namely by not extrapolating the data available to energies, where we have no evidence for it, but by approaching the question heads on by testing how far one can extrapolate. Take for instance the number of particle generations in the standard model. Three generations have been observed amounting to 12 matter particles. Could there be any more particles, a fourth generation, we have not yet observed? Requiring future theories to have only three generations may put too strong of a constraint on theory space. However, researchers have recently combined experimental data on Higgs searches from the particle accelerators LHC and Tevatron to conclude that a fourth generation of the standard model can be excluded with 5 σ. Similarly, take the example of the speed of light. It has been measured using a range of different methods, from astronomical measurements to interferometry. So its value is well tested and so for any future theory, one may require it to have that fixed constant value. However, do we really have good reason to believe that it is actually constant? Some physicists have suggested the possibility of theories with a varying speed of light (Magueijo, 2003). There has been, however, controversy about the viability of these theories. As Ellis and Uzan (2005, p.12) point out in arguments against these variable speed of light theories: "[t]he emphasis must be put on what can be measured", which leads to the constraint that "only the variation of dimensionless quantities makes sense". So we see that a seemingly simple concept like the speed of light is actually "complex and has many facets. These different facets have to be distinguished if we wish to construct a theory in which the speed of light is allowed to vary" [ibid.]. These examples illustrate that the topic of how data can constrain theory space is a highly non-trivial matter that needs much more careful analysis.

So we have illustrated how the different problems could in principle be addressed. While each of these problems need to be discussed in much more detail, we can already recognise some more general features for scientific methodology. Most importantly, it requires a careful reorientation in scientific practice. When I cannot test the novel predictions of the developed theory, I have no other choice but to assess carefully the elements that led me to it. This then may allows a confirmatory assessment in terms of theory space assessment. It is actually not true, that there is no other choice. Alternatively, one can just wait and hope that further progress in technology will soon catch up with the energy scale, where the novel empirical predictions are being made. The trouble with this second possibility is that many of those theories where non-empirical theory assessment is necessary, i.e. theories of quantum gravity, are by construct theories at the Planck scale and it is unclear whether in the foreseeable future experiments will be able to probe these scales. Non-empirical theory assessment by trying to open up a new option for assessing theories in the absence of empirical data, opens up a whole set of new questions regarding scientific methodology that future work needs to address.

2.5 Conclusion

In modern fundamental physics we have many instances where we are lacking empirical evidence. For instance, theories of quantum gravity have been developed in the absence of direct empirical evidence for decades now. In most of the twentieth century, it did not matter how theories were developed. After all, once we developed the theory we could just test it. Scientific practice itself was not affected so much by scientific methodology. As long as experiments can be done there is enough external guidance, that the developed methodologies did not need to affect the practice much. But if experiments are lacking, a whole new approach of theory development and assessment needs to be implemented.

We have shown how one can understand Dawidian non-empirical theory assessment as a conservative extension of empirical theory assessment. The method is based on the idea that one can assess theories based on an assessment of how constrained theory space is. When assessing theory space there are many ways we could have mistakenly constrained theory space. It starts with the empirical data and theoretical constraints we take as given. This is followed with the very scientific problem one aims to solve up to the background mathematical structure one assumes. The many misidentifications of constraints that need to be satisfied can lead to a false assessment of theory space. By identifying the very problems that can lead to a mistaken assessment, we have identified the elements that need to be addressed to guarantee a reliable assessment of theory space. The reliability of non-empirical theory assessment depends crucially on an accurate assessment of theory space.

One may argue that these problems just show that it is impossible to assess theories non-empirically. This might very well be the way it turns out, but it does not have to. Just as Kyle Stanford argued by using the history as evidence against a realistic interpretation of current theories, one can use it as an argument for the fact that scientists did after all find theories, which were previously unconceived. Now this happened, when scientists were not actually aiming at systematically exploring theory space. What if they were? What if scientists would actually explore theory space more systematically. There is of course the possibility that there are many possible alternatives that are currently not excluded and in those cases there will be no non-empirical support for any one particular. But it also may be the case that a wide range of theory space can be excluded and that these suggest that there really cannot be any alternatives. In those cases one may use this to confirm the theory.

Let us end on an important point. Nothing in the above suggested in any way that non-empirical theory assessment is supposed to replace empirical investigation. Ideally, it complements empirical confirmation (Fig. 2.4). As we outlined in Sect. 2.2.2 we confirm a theory first and foremost by empirical data. This is the basis from which non-empirical theory assessment sets off. It addresses the question of how far the available empirical data reaches within theory space. It is this point, which is addressed in non-empirical theory assessment and therefore can only complement empirical confirmation and cannot replace it. Non-empirical

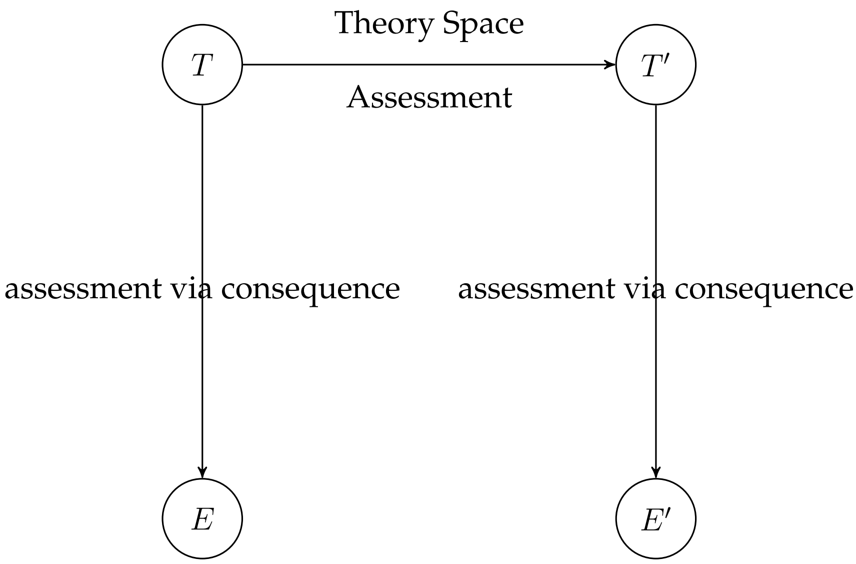

FIGURE 2.4: Empirical and non-empirical assessment complement each other.

theory assessment tries to make the most out of the empirical data we have.

Chapter 3

The No Alternatives Argument and Theories of Quantum Gravity

3.1 Introduction

Research in fundamental physics has changed in the last three to four decades. Theoretical work in, e.g. String Theory or other theories of quantum gravity, has been developed independently of experiment and does not begin with observed phenomena in need of an explanation. While, historically, it has often been the case that theory precedes experiment, in many of the current instances no empirical tests of the proposed theories are possible in the foreseeable future. Nevertheless, scientists trust these theories even in the absence of empirical data. Are scientists justified in trusting these theories? And if yes, how do they assess these theories in the absence of empirical data?

In a recent book Richard Dawid (2013) addresses these questions.The reason, he argues, is based on the idea that one can assess theories non-empirically by assessing the extent to which their domain is scientifically underdetermined. Roughly speaking, the more limited scientific underdetermination in a particular case is, the more likely it is that the considered theory is the right one. But how can we gain access or recognise whether scientific underdetermination is limited? While Dawid does not consider a direct assessment by exploring theory space, he considers an indirect approach via three observations that, as he argues, allow to argue for limitations on scientific underdetermination. Let us consider each of these in the context of String Theory, where empirical confirmation is currently not possible.

The first argument is the No Alternatives Argument (NAA), which is based on the observation that there are no viable alternatives to String Theory that can give a unified account of elementary particle interactions and gravity. The second argument is the unexpected explanatory coherence argument (UEA), which is based on the observation that string theory has led to several surprising results which were not to be expected and which lead to a more coherent overall picture. His last argument is the meta-inductive argument (MIA) from the success of other theories within the research program to the currently assessed one. The idea is that string theory is part of a long research program in high energy physics within which e.g. the Standard Model already had no alternatives, led to a consistent unified description of the nuclear forces, and has turned out to be highly successful. It is the conjunction of these three arguments which together establish the non-empirical evidence for String Theory.

What is meant by *non-empirical* evidence is not that it is not based on observations but that the evidence does not fall within the domain of applicability of the theory. That is, the theory by itself does not entail or make it more likely that there are no alternatives to it. These arguments allow us, so he argues, to assess scientific theories non-empirically. Supporting this, Dawid, Hartmann, and Sprenger (2015) have recently showed that the non-empirical assessment of scientific theories based on the NAA can be considered confirmatory in a Bayesian sense. This is a remarkable result since it explicitly shows that scientific confirmation is not necessarily restricted to the empirical realm of the theory. It is this confirmatory result we will be focusing on.

We will start in Sect. 3.2.1 by briefly reviewing the NAA as presented in (Dawid, Hartmann, and Sprenger, 2015). In Sect. 3.2.2 we analyse the definition of non-empirical evidence in the NAA as presented in (Dawid, 2013) and argue that it is inadequate if intended to be applicable to cases like String theory. We then propose a different, more adequate problem-oriented definition, which allows its application to the intended theories (Sect. 3.3). In Sect. 3.4 we conclude from the analysis before that either the No Alternatives Argument trivialises to the extent that everyone is justified to work on what they work on, or it commits one to seek for independent meta-inductive support regarding one's problem set, which in the cases most needed, i.e. theories of quantum gravity, cannot be sufficiently established.

3.2 The No Alternatives Argument

Before we detail the argument, it might be useful to consider the reasoning behind the NAA in a setting where it may be more familiar but different from the scientific theory context. Consider a sick patient going to a doctor. The doctor based on the statement of the patient, in addition to certain tests, collects a set of symptoms. She thinks about the possible diseases that could give rise to the set of symptoms and comes up, after having thought about it sufficiently, with only one disease which would account for all the symptoms. There might, of course, be diseases, which have not yet been discovered and which would equally well account for the symptoms. She is, however, confident that these specific symptoms have been researched sufficiently. Based on this reasoning she treats the patient accordingly. This quite reasonable case is the NAA applied to a medical setting, where symptoms correspond to the set of constraints and the disease corresponds to the theory. We will come back to this analogy in later sections where it will be helpful in evaluating the advantages and disadvantages of the NAA.

3.2.1 The Argument

To establish that non-empirical evidence can confirm a theory in a Bayesian sense[1] Dawid, Hartmann, and Sprenger (2015) aim to show that

$$P(T|F_A) > P(T), \tag{3.1}$$

where T stands for the proposition that the hypothesis H is empirically adequate, and F_A stands for the non-empirical evidence that the scientific community has not yet found an alternative to hypothesis H that fulfills a set of theoretical constraints, $\mathcal{C}$, explains the existing data, $\mathcal{D}$, and gives distinguishable predictions for the outcome of some set, $\mathcal{E}$, of future experiments. Compare this to the situation where some phenomenon that has been predicted by a theory is observed. The observation of that phenomenon is considered empirical evidence and is a direct consequence of the theory. However, since F_A is not a consequence of the theory, i.e.

[1] See e.g. (Bovens and Hartmann, 2004) or (Hartmann and Sprenger, 2010) for treatments of Bayesian confirmation theory.

it is non-empirical, its relation to T cannot be direct. It simply does not follow from the adequacy of the hypothesis that there are no alternatives. Dawid et al. now use the following interesting observation: while F_A and T are not directly linked, the connection between them may be mediated via a third random variable. And it may be the case that through that mediating variable, F_A will be able to confirm T. The candidate proposed between F_A and T is the proposition Y_k, which states

Y_k**:** There are k adequate and distinct alternatives, which satisfy a set of theoretical constraints, $\mathcal{C}$, are consistent with the existing data, $\mathcal{D}$, and give distinguishable predictions for the outcome of some set, $\mathcal{E}$, of future experiments.

Y_k could offer a link between F_A and T because the existence of a low number of alternatives would explain why scientists have not yet found any alternatives to it and therefore explain F_A. At the same time, if there are not many alternatives to one's theory, one may argue that this makes it more likely that the hypothesis one already has is empirically adequate. Together with certain assumptions[2] on the relation of F_A, T and Y_k and the corresponding probabilities, Dawid et al. prove that, indeed, $P(T|F_A) > P(T)$, and therefore F_A confirms the empirical adequacy of H.

There is, however, a caveat, which Dawid et al. address: the reason why scientists have not yet found an alternative could simply be due to the difficulty of the problem posed rather than the lack of alternatives. So there could very well be many alternatives, all of which are extremely difficult to find. Knowledge of F_A then only confirms the difficulty of the problem rather than the theory itself. Neither this problem nor the

[2]The assumptions, although obviously crucial for the proof, are not further discussed here. For the purpose of this paper it is assumed, with the authors, that they are reasonable. See (Herzberg, 2014) for a recent discussion of the assumptions.

reasonableness of the assumptions necessary to derive the theorem will play a role in this paper. It is rather the question: Is the non-empirical evidence F_A the correct evidence, to put limitations on scientific underdetermination? Let us start by stating it explicitly:

F_A: The scientific community has not yet found an alternative to hypothesis H that fulfills a set of theoretical constraints, $\mathcal{C}$, explains the existing data, $\mathcal{D}$, and gives distinguishable predictions for the outcome of some set, $\mathcal{E}$, of future experiments.

Whether F_A can be observed depends on whether $\mathcal{C}$, $\mathcal{D}$ and $\mathcal{E}$ can be defined sufficiently precisely and whether they are actually the right constraints. This needs to be addressed before the scientific community can possibly come to an agreement regarding the non-existence of alternatives. We will now turn to an analysis of $\mathcal{C}$, $\mathcal{D}$ and $\mathcal{E}$ as presented in (Dawid, 2013), before discussing its inadequacy and its appropriate reformulation in Sect. 3.3.

3.2.2 Dawid on $\mathcal{C}$, $\mathcal{D}$ and $\mathcal{E}$

How can we establish non-empirical evidence of the F_A kind? Dawid et al. say: find a theory that "fulfills a set of theoretical constraints, $\mathcal{C}$, explains the existing data, $\mathcal{D}$, and gives distinguishable predictions for the outcomes of some set, $\mathcal{E}$, of future experiments". If, after considerable effort, no alternative to that theory has been found that is able to fulfill $\mathcal{C}$, $\mathcal{D}$ and $\mathcal{E}$, then F_A has been established and the NAA can be applied. So it is $\mathcal{C}$, $\mathcal{D}$ and $\mathcal{E}$ that put constraints on theory space and which lead, in the ideal case for the NAA, to only one theory able to satisfy them. Dawid, Hartmann, and Sprenger (2015) do not discuss the constraints and what they are supposed to stand for any further. Dawid (2013) considers a changed F_A corresponding to the statement "that scientists have not found any alternatives to theory H which are (expected to be) consistent with the available data" but does implicitly address the above mentioned constraints in other parts of his book. The analysis below of $\mathcal{C}$, $\mathcal{D}$ and $\mathcal{E}$ is therefore an elaboration of (Dawid, 2013) in the context of their confirmatory result in (Dawid, Hartmann, and Sprenger, 2015).

Let us consider first the set of empirical data $\mathcal{D}$. The precise statement of what the role of the set $\mathcal{D}$ for the non-empirical evidence F_A is, is quite important and has been used in different ways by Dawid, Hartmann, and Sprenger (2015). They sometimes use the statement that theory H "explains empirical data $\mathcal{D}$". They are more careful, however, in using the phrase "being consistent with existing data $\mathcal{D}$" when formulating Y_k. That this difference in formulation is important becomes clear in cases where one most crucially wants to apply the NAA to, namely for theories of quantum gravity. If applied to, for instance, String Theory, the NAA cannot be applied if one requires an explanation of the empirical data.[3] String Theory does not yet offer an explanation of $\mathcal{D}$ and therefore, strictly speaking, the NAA cannot be applied to it. The weaker form, namely the requirement that the theory is consistent with the data $\mathcal{D}$ does allow one to use the NAA, since String Theory is, as far as one knows, consistent with the standard model of particle physics. Although "being consistent with data $\mathcal{D}$" is a weaker statement than "explaining $\mathcal{D}$", it still has some strength. If, for instance, the theory turns out not to contain fermionic particles, the theory would fail to possibly account for the matter content of the universe. This was actually the case for bosonic string theory, which was rejected due to this problem. Although, "being consistent with data $\mathcal{D}$" is more adequate for the application of the NAA to String Theory, it might not be adequate if the hypothesis was actually

[3] "Explanation" should here be understood in the sense that the theory is able to account for the empirical data, i.e. to either predict or retrodict the data.

introduced to account for some observed phenomenon, say the hydrogen spectrum or dark matter signatures. One still would want to apply the NAA in these cases but then "being consistent with data $\mathcal{D}$" would not suffice. This already suggests the problem that the precise statement of what H needs to accomplish with respect to $\mathcal{D}$ may strongly depend on what the theory actually aims to accomplish in general. We will get back to this later.

Let us now turn to the set $\mathcal{C}$ of theoretical constraints that need to be satisfied by a theory. These correspond to what are called scientificality conditions in (Dawid, 2013).[4] Scientificality conditions set the framework within which claims regarding the number of alternatives can be made. Given the available data the underdetermination[5] of theories considered are then constrained by ampliative rules of theory development which are the scientificality conditions. Following (Laudan, 1996) examples would be the principle of induction, the exclusion of ad-hoc explanations for individual events, simplicity assumptions, etc. However, Dawid argues that "a precise specification of scientificality conditions is not necessary" (Dawid, 2013, p.61). Given the predictive success in the research program, it can be assumed that the scientificality conditions are stable enough to assess limitations on scientific underdetermination. So to sum up, it is the available data $\mathcal{D}$ together with the scientificality conditions $\mathcal{C}$ which set the framework within which claims regarding underdetermination can be evaluated.

Finally, let us now consider the set $\mathcal{E}$ of future experiments for which

[4] This has been pointed out to me by Richard Dawid.

[5] Dawid uses the notion of scientific underdetermination, which might come closest to the one considered by scientists in theory development. It has similarities to the notion used in (Sklar, 1975) and (Stanford, 2001; Stanford, 2006). The differences are not crucial for what follows. See (Dawid, 2013, pp. 46-47) for details.

the theory should be able to make distinguishable predictions for its outcomes. This completes the constraints in the sense that it sets the energy scale at which the question regarding the number of alternatives is phrased. That is, the question of how many alternative theories have been found that are consistent with the empirical data $\mathcal{D}$, are established based on scientificality conditions $\mathcal{C}$, and give a coherent picture at the energy scale defined by $\mathcal{E}$.

3.3 A Problem-oriented Reformulation of F_A

3.3.1 Problems with Dawid's Specification of $\mathcal{C}$, $\mathcal{D}$ and $\mathcal{E}$

We have now presented the constraints $\mathcal{C}$, $\mathcal{D}$ and $\mathcal{E}$ as defended by Dawid (2013). Even allowing for the flexibility in whether the data needs to be explained by the theory or whether the theory needs to be consistent with the data only, the question remains whether the specification of $\mathcal{D}$ and $\mathcal{E}$ is sufficiently precise to establish F_A in such a way that there will be agreement among scientists? There are several reasons why this is problematic.

Consider first the following statement by Dawid which intends to strengthen the NAA in the context of String Theory: "[String] theory is the only viable option for constructing a unified theory of elementary particle interactions and gravity" (Dawid, 2013, p. 31). This statement differs significantly from the no alternative claim above. The correct observation supporting the NAA in the sense outlined above would state that "String Theory is the only theory at the Planck scale ($\mathcal{E}$), that is consistent with the available empirical data $\mathcal{D}$". This statement differs significantly from the above in the sense that unification is not required. That is, unification, although a crucial motivation for String Theory proponents, is not part of Dawid's constraints. If we do not require unification, however, one should take seriously positions like that of Weinberg (2009) and Donoghue (1994) that defend an effective theory approach to quantum gravity, which lacks the unificatory feature of String Theory.

Second, empirical data $\mathcal{D}$ does not necessarily determine the energy scale $\mathcal{E}$ at which we expect a new theory. Consider, for instance the need for dark matter particles. The scale at which the theory of dark matter has to be to account for the data is not uniquely fixed. There are proposals for dark matter candidates ranging several orders of magnitude in energy scale. According to the constraints outlined above, we can apply at each of those energy scales the NAA, since there might be no alternative to that specific theory of dark matter at that scale. This, however, corresponds to the wrong comparison class. Our trust regarding the specific dark matter candidate relies on the number of alternative theories able to give rise to the observed implications of dark matter. However, the observations themselves, e.g. galaxy rotation curves, do not set the energy scale at which dark matter candidates are to be expected.

Third, if we want the application of the NAA to depend that strongly on the energy scale, we have to rely on our current theories to tell us where interesting new effects are to be expected. In terms of the standard model of particle physics, one might argue, for instance, for a new theory at the grand unified scale (i.e. 10^{16} GeV). If we add gravitational interactions, the Planck scale at 10^{19} GeV is the scale at which we expect a new theory. But there are two problems with this. First, the specific energies will depend on many unconfirmed assumptions. For instance, the precise scale of grand unified theories will be different if we only assume the standard model, compared to assuming that supersymmetry plays a role at the TeV-scale. Second, the number of scales where one would be motivated to look for theories guided by energy scales alone is very limited and is therefore not sufficient to account for the many instances, where the NAA is supposedly applied to.

The above discussion should suffice to justify considering the number of alternatives not solely with respect to $\mathcal{C}$, $\mathcal{D}$ and $\mathcal{E}$. These constraints are simply not sufficient to apply the NAA to the many instances it ought to be applied to. In what follows, we will argue that the *scientific problems* themselves should be the focus. The set of alternative theories that need to be compared are determined with respect to a set of problems that they ought to solve.

3.3.2 Embedding Dawidian Assessement into a Laudian Framework

The NAA is based on the observation of scientists regarding the number of alternative theories that are able to do something. If we want to apply the NAA, we have to determine what this "something" is. As argued in the last section, $\mathcal{C}$, $\mathcal{D}$ and $\mathcal{E}$ are not sufficient, to determine this in most cases. I will, following Laudan, take the aim of scientific theories to be that of providers of solutions to scientific problems. The NAA is then based on the observation of scientists regarding the number of alternative theories that are able to provide solutions to a given set of problems. I will not defend here that this is the only viable approach, but consider Laudan to be convincing when he claims that the "view of science as a problem-solving system holds out more hope of capturing what is most characteristic about science than any alternative framework has" (Laudan, 1978, p.12).

According to Laudan there are two kinds of scientific problems: empirical problems and conceptual problems. Both of these, empirical problems to a lesser degree, "have no existence independent of the theories which exhibit them" (Laudan, 1978, p.48). So whether a theory T' solves a problem P will depend on theory T for which the problem P arises. So to assess a theory T' non-empirically, we will have to embed it within, what Laudan calls, its research tradition. To accommodate this embedment of Dawidian theory assessment with research traditions we need to complement the above constraints. We will understand scientific problems within our account as the set of constraints that we require the future theory to satisfy.

The first constraint we consider are theoretical constraints $\mathcal{T}$, which are determined by the current confirmed theories. For example, any future theory of high energy physics is constrained by requirements of Lorentz invariance and the $SU(3) \times SU(2) \times U(1)$ gauge group of the standard model for those domains already tested. These constraints can be quite severe. For instance, any future theory should contain the $SU(3) \times SU(2) \times U(1)$ gauge group as a subgroup in order for it to be able to account for the previous successes. This put strong constraints on the number of alternatives to H.

However, we mentioned above that unification is a crucial motivating force in accepting theories like String Theory. It is the aim to develop theories that are unificatory. But where does it fit in the above framework? One might argue that it is a theoretical constraint and should be part of the set $\mathcal{T}$. However, the status of theoretical constraints like Lorentz invariance and the standard model gauge group is of a different character then the unification for all fundamental forces. The former are already part of theories which have been experimentally confirmed. The unification of all fundamental forces, however, is part of a methodological approach within high energy physics which has been successful in the past.

That unification is not a necessary assumption has been argued for by philosophers of science like Nancy Cartwright. She argues that "[t]he laws that describe this world are a patchwork, not a pyramid." (Cartwright, 1999, p. 1) and that each theory within their respective domains may be simultaneously true without the need to unify all theories within one overarching theory. She is thereby denying the very assumptions motivating string theorists. The point here is not to defend Cartwright vs. the unificationists but to show that the assumption underlying a unificationist position is based on a different and less empirically justified basis than those theoretical constraints which have been successful ingredients of experimentally confirmed theories. It might, therefore, be useful to distinguish between a set $\mathcal{T}$ of *theoretical constraints* which are strong theoretical principles within the research program justified by empirical evidence and a set $\mathcal{T}'$ of *theoretical assumptions* which are not justified by the evidence available. In fact, there is empirical evidence against unification in the sense of grand unified theories (which unify the strong and the electroweak theory). The coupling constants of the different forces do not meet at higher energies and so, strictly speaking, point against gauge coupling unification of the standard model forces. Further mechanisms have to be introduced before a meeting of the gauge coupling constants can be achieved. One might object that, given the shaky ground on which theoretical assumptions such as these lie, they should not be used to put constraints on the set of possible theories. We will later get back to this legitimate objection, but note that in practice they do play a fundamental role.

The distinction between theoretical constraints and theoretical assumptions is not solely introduced for the purpose of making a distinction with respect to unification, but is applicable to a wide array of cases in which one grounds a theory on additional assumptions which are not empirically justified. To illustrate this let us consider another example, the hierarchy problem. The hierarchy problem is the problem that there is a huge gap between the electroweak scale and the scale at which gravity becomes important for particle physics, namely the Planck scale. This seemingly aesthetic problem has wide ranging consequences if applied to the quantum field theory of scalar particles. The self-energy of scalar fields like the Higgs particle as calculated in quantum field theory leads to (quadratic) divergencies which, however, can be accounted for through renormalisation. In the renormalisation procedure the divergencies can be 'cut' at the next scale and then subtracted from the so-called bare mass. If nothing happens in between, the next scale is the Planck scale and the corresponding subtraction would be possible only if one fine-tunes the corresponding bare mass to recover the physical Higgs mass at the relatively low tera-electronvolt scale. Since the 'distance' between the electroweak scale and the Planck scale is huge, the amount of fine-tuning needed is huge as well. Fine-tuning is highly unnatural, or so the physicists claim, and therefore is in need of an explanation[6]. The wish for a non-fine-tuned Higgs mass should, so at least I argue, be considered part of the theoretical assumptions $\mathcal{T}'$ rather than part of the constraints $\mathcal{T}$, since the standard model has all the *mathematical* tools necessary to account for the TeV scale Higgs mass.

[6]This led physicists even as far as to propose the existence of large extra-dimensions (Arkani-Hamed, Dimopoulos, and Dvali, 1998) or warped extra-dimensions (Randall and Sundrum, 1999) to solve the problem.

Now that we have complemented $\mathcal{D}$ and $\mathcal{E}$ with $\mathcal{T}$ and $\mathcal{T}'$ we can reformulate the proposition F_A. The aim of a theory may be to account for some empirical data from $\mathcal{D}$ while satisfying certain theoretical constraints $\mathcal{T}$. While another theory may aim to provide a solution to some theoretical assumption $\mathcal{T}'$ which needs to be satisfied, while at this point only tries to be consistent with the empirical data $\mathcal{D}$. So, while there is no doubt that the constraints all play an important role in theory development, they may not all simultaneously represent the constraints scientists consider when confronted with a problem. Some may rather represent long-term hoped-for achievements but are not crucial at the first step in offering a solution to a problem. So the problem or set of problems, call it P, should be the focus of the NAA. So P takes elements from $\mathcal{T}$, $\mathcal{T}'$, $\mathcal{D}$ and $\mathcal{E}$ (but usually does not need to exhaust all of $\mathcal{T}$, $\mathcal{T}'$, $\mathcal{D}$ and $\mathcal{E}$) and specifies what, with respect to these constraints, needs to be achieved.[7] This leads to the following reformulation of F_A:

F_A^P: The scientific community has not yet found an alternative to hypothesis H that fulfills the constraints determined by the problem set P.

So the NAA should be understood as being always an argument with respect to a specific problem set P. Let us now turn to specific problem sets and focus on the case of theories of quantum gravity.

[7]The scientificality conditions $\mathcal{C}$ will remain in the background, setting the methodological framework within which theories are being developed.

3.4 The NAA and Confirming Theories of Quantum Gravity

3.4.1 Finding P

Having re-situated the NAA within a problem-oriented approach, the remaining source of danger for an adequate account of non-empirical evidence of the NAA kind is the problem set itself. So how should one determine the problem set P? A problem set, which is based on false constraints will put inappropriate limitations on scientific underdetermination, thereby incorrectly leading to a not well justified confirmation of a theory via the NAA. As we saw, when we discussed the constraints $\mathcal{T}$, $\mathcal{T}'$, $\mathcal{D}$ and $\mathcal{E}$, some elements of a problem set may either lack or be less epistemically justified than others, necessitating the distinction we made between theoretical assumptions $\mathcal{T}'$ and theoretical constraints $\mathcal{T}$. While, for instance, any future theory of particle physics is constrained by the effective group structure at low energies (part of the theoretical constraints $\mathcal{T}$, which everybody agrees upon), they are also constrained by the desire that they should be given by a simple unified group (part of the theoretical assumptions $\mathcal{T}'$, not unanimously agreed upon). Given the trouble of finding the right problem set in cases where theoretical assumptions are involved, we will now propose two possible ways to move forward. First, since the elements of $\mathcal{T}'$ are beyond the empirically confirmed part of the theories within the research program, any problem set P proposed by scientists will do. Second, even though the elements of $\mathcal{T}'$ are not empirically justified, they may be justified non-empirically. Let us consider each case now in more detail.

If there is no way to epistemically justify the problem set, any scientist may choose her own preferred set of problems. It is reasonable to assume that scientists within their own research tradition will have a preferred set of problems. With regard to that specific problem set, scientists will agree on the number of alternatives, even though they might disagree on the specific problem set itself. There is of course a danger of trivialisation. It seems quite plausible to assume that for any theory of your liking you can find a set of constraints, such that there will be no other theory able to satisfy them. So since with respect to this set of constraints, there will be no alternative theories, we may want to apply the NAA to confirm our theory. It seems obvious, however, that we do not accept this as being confirmatory, for the simple reason that all theories would then be confirmed.

So does the argument trivialise in the less extreme case? Not completely. Scientists work on the specific theories they are working on, because they consider the theory they use to be the most appropriate in light of the set of problems they wish to address, even though the set of problems is not itself epistemically justified. If many alternatives were able to address the same problem set, their trust in their specific approach may decrease. The confirmatory result that follows from the NAA should in this case not be understood as confirmatory, but as providing an explanation for why scientists work on the theory they use, given their specific problem set.

Let us consider the corresponding case in the medical diagnosis setting. Having one specific set of symptoms all doctors will, ideally, agree that there is only one disease able to account for all the symptoms. However, they might not agree on the set of symptoms. The patient's paediatrician might deviate from that set of symptoms given specific knowledge she obtained during the patient's childhood. Maybe one doctor thinks the patient is a hypochondriac[8], suggesting that the set of symptoms is unreliable. The patient's cardiologists may have a different perspective regarding the patient's explanations, which again may give rise to a different set of symptoms, etc. So, although, keeping the set of symptoms fixed all doctors will likely agree on the disease, there might be disagreement regarding the patient's set of symptoms. However, each treatment the doctors suggest can be understood with regard to their specific set of symptoms, just as one can understand the scientist's commitment given her specific problem set.

Let us now turn to the more interesting case where one specific problem set is justified over and above the others. In this case the NAA by itself may provide theory confirmation. But how can we justify one specific problem set rather than another? Let us consider more concrete cases. Why is a theoretical assumption of unification or naturalness justified? Since these constraints are theoretical assumptions and therefore not at this point empirically justifiable, one has to consider the appropriateness of these assumptions within the bigger research tradition. Their value then may be determined by their success within the research program as a whole. So one is committed to meta-inductive arguments[9] regarding past successes of instances of theoretical assumptions. This again is a case of non-empirical evidence.

[8]To take the analogy a little further: there is no direct test to see whether a patient is a hypochondriac or not. So if hypochondria plays a role in ones judgement of the disease, hypochondria comes close to what we called a theoretical assumption. One can provide support for hypochondria by considering the patient's history (i.e. research tradition in the scientific context). This is the route we will consider next.

[9]The meta-inductive argument employed here is, however, crucially different to that employed by (Dawid, 2013). His meta-inductive argument considers previously successful instances of the NAA within the research program, while the meta-inductive argument here considers previously successful instances of a theoretical assumption within the research program.

How can one meta-inductively support the inclusion of a theoretical assumption in the problem set? Let us consider the case of unification. The theoretical assumption of unification has been a successful assumption within the history of particle physics. To aim for unification provides a methodological approach which has led to many empirically successful unifications which in turn led to novel empirical predictions that were confirmed. The unification of electric and magnetic phenomena, and later the unification of electromagnetic and weak interactions are examples. So there seems to be historical evidence that the theoretical assumption of unification has been within the research program the right one to use. I.e. there is meta-inductive support for the inclusion of the theoretical assumption of unification in the problem set.

There is, however, a problem with meta-inductive arguments of the sort mentioned above. How are we supposed to compare instances of unification in history? Is the unification of planetary motion and falling apples within a gravitational theory of the same kind as the unification of electric and magnetic phenomenon or the unification in the electroweak theory? Are unifications proposed in grand unified theories comparable to any previous unification before? Do we have unification only if it is provided by a unification within a simple mathematical group?[10] If yes, is the case of electroweak interactions a different kind of unification? Unless there is a clear definition of what is meant with unification it is difficult to compare historical instances in a precise way. However, after having defined the relevant notion precisely and having made a case for why these are commensurable, will there still remain enough evidence for the purposes of the meta-inductive argument? And even then, what

[10]Compare for instance the unification achieved by (Pati and Salam, 1974) ($SU(4) \times SU(2)_L \times SU(2)_R$) to that by Georgi and Glashow, 1974 ($SU(5)$).

does "enough" mean? These are difficult questions that need to be answered before one may use them in providing support for including the theoretical assumptions $\mathcal{T}'$ in the problem set P. But let us for the moment ignore this problem and actually consider it solved. We will turn to quantum gravity now, where even if we allow for the meta-inductive support of theoretical assumptions, a problem remains.

3.4.2 The Problem Set P and Quantum Gravity

There are many different accounts and approaches to quantum gravity. There are String Theory, Loop Quantum Gravity (LQG), Causal Dynamical Triangulation, Canonical Quantum Gravity, Asymptotic Safety and several others. They differ significantly in the amount of research and "success" they have had. The two largest communities within quantum gravity research are the defenders of String theory and Loop Quantum Gravity. We will be focussing on these two.

In a recent review of Richard Dawid's book, the physicist Lee Smolin (2014) applies the NAA to the case of LQG. The way he does it, amounts within our treatment to an appropriate reformulation of the problem set. He says:

> "[T]here is no alternative to [LQG] as a successful solution to the problem of giving a mathematically consistent and ultraviolet finite quantisation of general relativity, in 3 + 1 spacetime dimensions, without extraneous assumptions. [...]. [I]t satisfies the principle of background independence – one of the principles of general relativity – that rivals such as perturbative quantum gravity and string theory fail to do. "(Smolin,

2014, p.1105) [11]

For the NAA to be reliable, the above mentioned features by Smolin, do not only need to be features of the theory, but should be constraints any theory of quantum gravity should satisfy. As an example, a string theorist would argue that the above constraint to 3+1 spacetime dimensions is too restrictive, as we cannot rule out the possibility of compactified extra dimensions. However, this is not the place to discuss all the required constraints any theory of quantum gravity needs to satisfy. It is actually extremely controversial what the set of constraints actually needs to be. We will instead focus on a more general problem we are confronted with, which already illustrates the difficulty in confirming any theory of quantum gravity at this point.

Let us consider one specific theoretical assumption, namely background independence[12]. Why is the theoretical assumption of background independence appropriately included within the problem set in the Loop Quantum Gravity program, while this may not be the case in String Theory? Let us assume that in LQG, the inclusion of the theoretical assumption of background independence within the problem set can be meta-inductively justified. LQG grew out of the project of quantising gravity, and most scientists began their careers working on Einstein's theory of general relativity. Background independence is therefore an

[11]Though tangential to the issues involved here, it is interesting to note how Smolin continues. He argues that the applicability of NAA to both String Theory as well LQG poses a problem for Dawid. He says "But if Dawid's criteria can be used equally well to support rival research programs, and to justify the attentions of two competing research communities, his argument must be judged to fail" (Smolin, 2014, p.1105). He then goes on to argue for empirical research as the only way to go. One does not need to go far to see why this specific criticism of non-empirical theory assessment is without any substance. He himself says a couple of paragraphs before: "The same experimental result can confirm two theories that contradict each other" [ibid.], which is of course completely right. However, following the logic of his criticism of Dawid, empirical arguments "must be judged to fail" as well.

[12]It is actually far from obvious what background-independence amounts to. See e.g. (Rickles, 2008; Belot, 2011; Pooley, forthcoming).

understandable and crucial assumption for any future theory within that research tradition. On the other hand, String theory grew historically out of particle physics.[13] Particle physicists are accustomed to the use of perturbative methods and so for them background independence did not necessarily play a crucial role in the development of the theory. The disagreement regarding the problem set may therefore be understood as a specific feature of theories of quantum gravity, namely that different research traditions meet in the quest for a theory of quantum gravity, each having different non-empirical evidence for their specific problem set. So at this point in history, non-empirical evidence has not accumulated enough evidential support to obtain a unique problem set, which would justify one theory over another.

The more substantive conclusion of the NAA, that is, its intended claim regarding theory confirmation, is strongly dependent on the strength and evaluability of the problem set. However, these may rely on meta-inductive arguments. The meta-inductive argument may be more easily applicable in some cases, like the Higgs mechanism, while it is so far not uniquely applicable in the case of theories of quantum gravity, since this is where different research traditions collide and so may not lead to a unique set of problems to solve.

Coming back to the medical diagnosis setting, we can have the situation, where all the expertise is brought together and the patient's history is considered in its entirety. This may allow for an agreement among the doctors for one specific set of symptoms for which it turns out there exists only one disease that accounts for all of them. So the NAA can be applied and the chosen disease be "confirmed". However, it might be the case that the different doctors with different specialities do not have

[13]See (Rickles, 2014) for an excellent recent history of string theory.

sufficient information provided by the patient or that the amount of information obtained through blood samples etc. are not sufficient to apply the analogue of the NAA at this point. Their various presuppositions will then justify the respective treatments which may vary from doctor to doctor. This is analogous to the case where one can apply the NAA both, as I argued above, to String Theory and Loop Quantum Gravity, where each may be justified in their respective problem sets. However, while the diagnosticians have to make a decision regarding their patient's health, there is no hurry in the physicist's decision to commit to a theory at this point. With time, more empirical, theoretical and even historical evidence may contribute to a convergence of the problem sets and an agreement among the scientists.

3.5 Conclusion

In this paper we have analysed the confirmation claim of the NAA by analysing whether the non-empirical evidence can be obtained in a way such that there is agreement among the scientists regarding the evidence. We have proposed a problem-oriented approach where a problem set does the work of placing limitations on scientific underdetermination with respect to a theory. The strong dependence on a specific problem set then leads us to consider two possible ways of determining the problem set. In the first, the problem set is simply chosen by an individual or a group of scientists, in which case anybody is justified by the NAA to work on whatever they are working on. In this case the NAA does not provide theory confirmation but accounts for scientific practice. In the second, one determines the problem set meta-inductively. However, where the history has not yet provided enough evidential support, as in theories of quantum gravity, this may currently not be possible. This may lead to the unfortunate consequence that the NAA may, in the cases where it is most needed (i.e. in theories where empirical evidence is missing) not yet be applicable, while in cases where one does have enough non-empirical support, empirical evidence can be given too (as in the case of the Higgs mechanism), so the NAA is not needed.

Chapter 4

The Epistemology of No-Go Theorems

4.1 Introduction

The brilliant physicist and mathematician John von Neumann claimed to have proven in his classic (Neumann, 1932) the impossibility to complete quantum mechanics by hidden variables. Thirty years later, Jauch and Piron (1963) still state that "[t]he question concerning the existence of such hidden variables received an early and rather decisive answer in the form of von Neumann's proof on the mathematical impossibility of such variables in quantum theory". Three years later Bell (1966) shows in his seminal work that "the formal proof of von Neumann does not justify his informal conclusion", saying later in an interview that "the von Neumann proof, if you actually come to grips with it falls apart in your hands! There is nothing to it. It's not just flawed, it's silly! [...] The proof of von Neumann is not merely false but foolish!"[1]. Thirty years later Mermin (1993), following Bell, still considers that "von Neumann's no-hidden variables proof was based on an assumption that can only be described as silly". Going forward in time another 17 years, Jeff Bub (2010, p. 1334)

[1]As cited in (Mermin, 1993, p. 88).

argues that "Bell's analysis misconstrues the nature of von Neumann's claim, and that von Neumann's argument actually establishes something important about hidden variables and quantum mechanics".

The details of the von Neumann no-go theorem will not concern us in this chapter, but this example of a history of a no-go theorem nicely illustrates the difficulty of interpreting no-go results in physics. Opinions about it varied between having established a "decisive answer" on the question of hidden variables to the proof being considered "foolish"; the whole debate now ranging more than eight decades. This is not to say that there was no progress or that there is not a way to understand the disagreement and its development. However, this example illustrates that the role of no-go theorems in physics differs significantly from the case of impossibility or no-go results in mathematics. When we prove something in mathematics, there usually does not seem to be that much disagreement about what the theorem means. This already hints at the more complex structure of no-go theorems in physics compared to those in mathematics. There is a plethora of examples in the history of physics where this more complex structure was not adequately recognised and where it was misunderstood what no-go theorems can imply. In this chapter we want to analyse abstractly the general implications one should draw from no-go theorems.

Talking about no-go theorems in general bears a danger. Are all no-go theorems similar enough to the extent that it makes sense to talk of no-go theorems in physics in all generality and what they imply? No! To analyse what no-go theorems imply one needs to study them case by case. However, being unaware of what the general role of no-go theorems can be bears the danger of misinterpreting what a particular result actually implies and can misdirect a whole research effort based on a misinterpretation of the situation. The aim of this chapter is to address this more general structure of no-go theorems and what the implications can be based on this more general structure.

I start in Sect. 4.2 with the presentation of a case study of a set of no-go theorems from particle physics. This provides us with enough detail to provide an analysis of the abstract argument structure of no-go theorems (Sect. 4.3). In Sect. 4.4 I discuss on the basis of the previous section each element of a no-go theorem in more detail. In Sect. 4.5, I consider the possible implications of no-go theorems more broadly. In particular, I argue that no-go theorems have a more complex structure than is usually assumed. This more complex structure, however, allows us to infer that no-go theorems cannot play the role of impossibility results in the strict sense and are best understood as 'go' theorems.

4.2 A Case Study: Combining Internal and External Symmetries

Our tactic in assessing what no-go theorems imply is to start by considering a specific historical example of a no-go theorem. More specifically it is an example of a set of no-go theorems, each aiming to establish the impossibility to combine internal and external symmetries.

The exemplar no-go theorem that we cover is a not much discussed episode in the history of particle physics. During the 1960s physicists tried to combine external and internal symmetries. External symmetries

are the symmetries of space and time. These can be the discrete symmetries of parity and time reversal or continuous symmetries like translations and boosts. The mathematical structure associated with the spacetime symmetry that physicists focused on in the 60s was the Poincaré symmetry, which contains the Lorentz symmetry and the symmetry under translation. One contrasts external symmetries with internal symmetries. Internal symmetries are symmetries of a corresponding internal 'space'. Examples are Gell-Mann and Ne'eman's $SU(3)$-flavour symmetry, which in modern terms, is a symmetry under the change of the flavour of quarks with respect to the strong force. Other popular examples are Heisenberg's $SU(2)$-Isospin of the neutron and proton or the standard model gauge group $SU(3) \times SU(2) \times U(1)$.
I will discuss two motivations for why physicists tried to combine internal and external symmetries (4.2.1). This will be followed by a discussion of some no-go theorems that culminated in the result of Coleman and Mandula in 1967 (4.2.2). Finally, I will discuss certain routes towards combining internal and external symmetries which were not affected by the no-go theorems (4.2.3).[2]

4.2.1 Why Combine Internal and External Symmetries?

Symmetries in physics are strongly related to the properties characterising the particles of the theory. The relation is outlined in the Appendix. To put it briefly: one looks for those operators that commute with the generators of the symmetry. The eigenvalues of these operators then correspond to the invariant properties of the particles. The properties thus related to the Poincaré group, i.e. the external symmetry, are spin and mass. For internal symmetries like SU(2) it is the isospin or for SU(3) it is

[2]See (Weinberg, 2011, Sect. 24) and (Di Stefano, 2000) for historical accounts and (Iorio, 2011) for a more systematic treatment, which we follow here.

the quark flavour. One can, of course, always combine internal and external symmetries trivially, by considering the direct product of the two groups. In this case, however, all elements of the internal and external group commute with each other and so remain independent. Therefore one is interested in the non-trivial combinations of the symmetry group. There were two main motivations behind the wish to combine internal and external symmetries for more details on these, which we now turn our attention to.

The Problem of Mass-splitting Consider the $SU(2)$-Isospin symmetry introduced in (Heisenberg, 1932). Although neutrons are neutrally charged and protons positively charged they interact equally under the strong force. This led Heisenberg to the $SU(2)$-Isospin symmetry which transforms between protons $|+\rangle$ and neutrons $|-\rangle$, i.e. $I_{\mp}|\pm\rangle = |\mp\rangle$ with $[I_+, I_-] = 2I_0$ and $I_0|\pm\rangle = \pm|\pm\rangle$.

The translation generator of the Poincaré group P_μ commutes with the $I_\pm$, i.e. $[P_\mu, I_\pm] = 0$. From this it follows that $P^\mu P_\mu|\pm\rangle = m^2|\pm\rangle$, where m is the mass of the states. That is, since the momentum generator commutes with the $SU(2)$ generator, the proton and neutron will have to have the same mass. Although this is a good approximation, protons and neutrons do not have the same mass. The idea now is, that a non-trivial commutation relation between them may lead to the known mass difference between the proton and the neutron. For instance, by assuming $[P_\mu, I_+] = c_\mu I_+$ one obtains after some manipulations using the changed commutation relations $P^2|+\rangle = I_+ P^2|-\rangle - c^2|+\rangle$ from which one can easily show $m_p^2 = m_n^2 - c^2$. One can then recover the hoped for mass difference by experimentally fixing the c^2 value. So by mixing internal and external symmetries the hope was to explain the mass difference of particles. This initial motivation turned out not to be significant, as nowadays

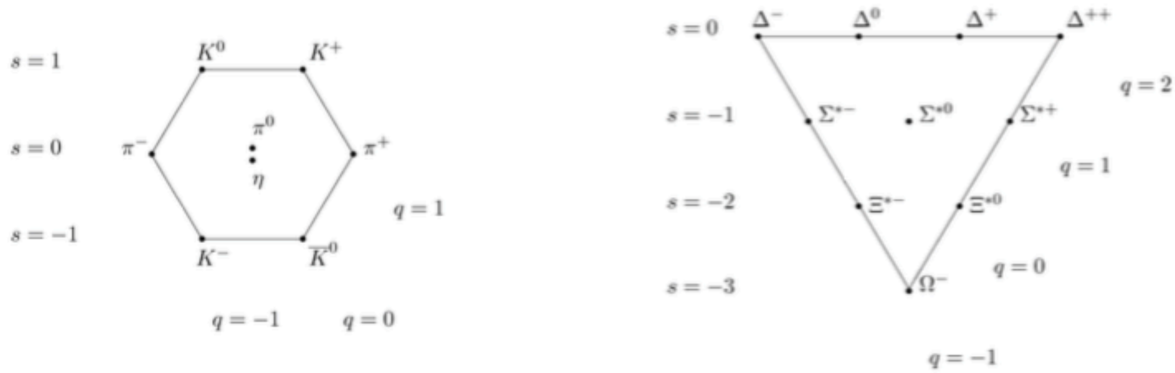

FIGURE 4.1: On the left is the spin-$\frac{1}{2}$ baryon octet and on the right is the spin-$\frac{3}{2}$ baryon decuplet, which both combine particles with different strangeness and charge. Source: Wikipedia

we know that protons and neutrons are composite particles made up of different quarks.

Unification The second motivation for combining internal and external symmetries is the methodological urge within the particle physics community to unify. If internal and external symmetries could be understood as following from one more general unified simple group, we would be one step further in the unification program within particle physics. Consider Gell-Mann (1964) and Ne'eman (1961)'s $SU(3)$-Flavour Symmetry. During the 1960s many new particles were being discovered and the relation between them was unknown. It was the SU(3)-flavour symmetry that allowed an understanding of the different baryons and mesons then discovered as elements within multiplets of the same group. There was for instance a baryon octet bringing together particles with different strangeness and charge but the same spin, namely spin-$\frac{1}{2}$, into one octet or a baryon decuplet similarly combining particles of the same spin, this time spin-$\frac{3}{2}$ particles, into a multiplet (See Figure 4.1).

Having unified particles with different strangeness and charge within multiplets the hope was to be able to unify particles with different spins within one multiplet as well. Since spin is an external property, this would amount to combining internal (strangeness, charge) and external (spin) properties. So bringing particles with different charges, strangeness and spins within a multiplet can only be achieved by bringing together internal and external degrees of freedom in a non-trivial way. One early step in this direction was the SU(6) symmetry group. The $SU(6)$ group was introduced and succeeded in unifying the baryon octet and decuplet into a 56-plet[3]. This gave rise to further attempts at unifying internal and external symmetries, since $SU(6)$ was not yet the end of the story. What $SU(6)$ achieved was a unification of $SU(3)$-flavour with non-relativistic $SU(2)$ spin. A full relativistic unification, i.e. one including the full Poincaré group, was then hoped for and attempted. But attempts failed, leading the way to several no-go theorems.

4.2.2 No-Go Theorems

Several no-go theorems were proposed between 1964 and 1967 culminating in the famous Coleman-Mandula theorem. The no-go theorems that were being developed ranged from mathematical to more and more physical arguments for the impossibility of combining internal and external symmetries. We will now mention three no-go theorems starting with the simplest argument made by McGlinn (1964) for the impossibility of combining internal and external symmetries.

In 1964 McGlinn, having the mass splitting problem from before in mind, proved the following theorem[4].

[3]See e.g. (Sakita, 1964) and (Gürsey, Pais, and Radicati, 1964).

[4]We follow O'Raifeartaigh's presentation of McGlinn's theorem in (O'Raifeartaigh, 1965) to allow for a more coherent nomenclature.

McGlinn Theorem: Let $\mathcal{L}$ be the Lie algebra of the Poincaré group, M and P the homogeneous and translation parts of $\mathcal{L}$, respectively, and $\mathcal{I}$ any semisimple internal symmetry algebra.

(a) If $\mathcal{T}$ is a Lie algebra whose basis consists of the basis of $\mathcal{L}$ and the basis of $\mathcal{I}$, and

(b) if $[\mathcal{I}, M] = 0$ (i.e. the internal symmetry is Lorentz invariant)

then $[\mathcal{I}, P] = 0$. Hence $\mathcal{T} = \mathcal{L} \times \mathcal{I}$.

So if (a) and (b) are satisfied, one can combine the internal group $\mathcal{I}$ with the external group $\mathcal{L}$ only trivially. Note this is a mathematical result, in the sense that it is not a result that follows from within the framework of a physical theory. As such it seems to be of a more general nature.

However, McGlinn's theorem gave rise to several papers which aimed to weaken the assumptions. For instance, early attempts by Michel (1965) and Sudarshan (1965) showed that to obtain McGlinn's result it is sufficient to assume that only one of the generators of the internal symmetry algebra $\mathcal{I}$ does not commute in (b). But it is especially assumption (a) that seems too stringent and unnecessary and which therefore motivated O'Raifeartaigh in 1965 to prove a more general theorem. Rather than building up the larger group starting from the Poincaré group, O'Raifeartaigh looked for the most general way to embed the Poincaré group into a larger group with the only restriction that the larger group is of finite order. The finite order of the larger group is necessary so that the so-called Levi decomposition theorem which forms the basis of his theorem can be applied. So with the only requirement that the group within which the Poincaré group is to be embedded be of finite order, O'Raifeartaigh was able to categorise the possible embeddings in the following theorem:

> **O'Raifeartaigh Theorem:** Let $\mathcal{L}$ be the Lie algebra of the Poincaré group, consisting of the homogeneous part M and the translation part P. Let $\mathcal{T}$ be any Lie algebra of finite order, with radical S and Levi factor G. If $\mathcal{L}$ is a subalgebra of $\mathcal{T}$, then only the following four cases occur:
>
> (1) $S = P$;
>
> (2) S Abelian, and contains P;
>
> (3) S non-Abelian, and contains P;
>
> (4) $S \cap P = \emptyset$.
>
> In all cases, $M \cap S = 0$.[5]

O'Raifeartaigh then goes on to discuss each possibility in detail. One thing that one can already see is that from a purely mathematical point of view it is possible for the internal and external symmetry to be combined in a non-trivial way. O'Raifeartaigh shows that case (1) reduces to the McGlinn case where one obtains $\mathcal{T} = \mathcal{L} \times \mathcal{I}$. Cases (2)-(4) are cases where the internal and external symmetries could possibly be combined non-trivially but these are, as O'Raifeartaigh argues, physically unreasonable. For instance, case (2) necessitates a translational algebra of more

[5]Some background may be helpful here: the Levi decomposition theorem states that any Lie algebra of finite order can be decomposed into the semi-direct sum of its radical (maximally solvable Lie algebra) and Levi factor (semisimple Lie algebra). Since P is abelian its first-derived algebra is empty and therefore solvable. M is semisimple therefore not solvable and contained in G. This leads to the four mentioned possible cases of decomposition.

than four dimensions, or case (3) has the problem that, due to Lie's theorem, any finite dimensional representation of a solvable non-abelian algebra has a basis such that all matrices have only zeros above the diagonal, i.e. are triangular matrices. This leads to the problem that one cannot always define hermitian conjugation. So unlike McGlinn's theorem, O'Raifeartaigh's theorem rules out a non-trivial combination of internal and external symmetries for physical reasons.

Although O'Raifeartaigh was able to generalise McGlinn's no-go theorem it was still considered to have shortcomings. One shortcoming was the need to consider only Lie algebras of finite order and the second shortcoming is the concentration on the one-particle spectrum only. Coleman and Mandula (1967) were able to account for both of these shortcomings by moving away from the mathematical framework of McGlinn and O'Raifeartaigh, towards a physical framework, namely S-Matrix theory, wherein the symmetries from before are the symmetries of the S-matrix.[6] This allowed them to consider n-particle spectra but still without the need to consider any specific quantum field theory. Also no need for finite order Lie algebras was necessary anymore. However, several physical and mathematical assumptions were introduced. The Coleman-Mandula Theorem states the following:

> **Coleman-Mandula Theorem:** Let $\mathcal{T}$ be a connected symmetry group of the S matrix, and let the following five conditions hold:
>
> 1. $\mathcal{T}$ contains a subgroup locally isomorphic to the Poincaré group $\mathcal{L}$;

[6]Coleman was already working on the problem of combining internal and external symmetries in 1965 when he was able to show that certain relativistic versions of $SU(6)$ had absurd consequences and should therefore be discarded (Coleman, 1965).

2. all particle types correspond to positive-energy representations of $\mathcal{L}$, and, for any finite mass M, there are only a finite number of particle types with mass less than M;
3. elastic-scattering amplitudes are analytic functions of the center of mass energy and of the momentum transfer in some neighbourhood of the physical region;
4. at almost all energies, any two plane waves scatter;
5. the generators of $\mathcal{T}$ are representable as integral operators in momentum space, with distributions for their kernels.

Then $\mathcal{T}$ is locally isomorphic to $\mathcal{L} \times \mathcal{I}$, the direct product of the Poincaré group and the internal symmetry group.

This represented the final blow to attempts in the community at unifying internal and external symmetries.[7] It is interesting to note that the physicists working on this unification project were actually hoping for the opposite result. While aiming for unification they apparently ended up showing its impossibility.

4.2.3 The Rise of Supersymmetry

As mentioned, the Coleman-Mandula theorem stopped much of the discussion on internal and external symmetries. The explicit assumptions above did not give rise to physicists attempting to weaken the assumptions, although some problems with them were known (see e.g. Sohnius, 1985). However, in the subsequent years, three different groups with

[7]With a single exception: Mirman (1969) made the more general claim that "the impossibility theorems have no physical relevance". This was followed by Cornwell (1971), where it is claimed that "Mirman's objections may be overcome without difficulty, and that the above-mentioned theorems do indeed relate to the physical situation".

completely different motivations were able to non-trivially combine internal and external symmetries.The first successful proposal was by Yuri Golfand and his student Evgeni Likhtman from the Physical Institute in Moscow.[8] The actual reason motivating Golfand to develop an extension of the Poincaré group is not clear. However, they try to account for parity violation in the weak interactions in their original paper. Although, they also state the following reason: "only a fraction of the interactions satisfying this requirement [i.e. being invariant under Poincaré transformations] is realised in nature. It is possible that these interactions, unlike others, have a higher degree of symmetry" (Golfand and Likhtman, 1971, p.323). So the search for this higher symmetry can be seen to have been their goal as well. Volkov and Akulov (1972) from the Kharkov Institute of Physics and Technology had other reasons for their development. They hoped to be able to describe the neutrino, then thought to be massless, as a Goldstone particle. Obtaining Goldstone particles with half-integer spin like the neutrino makes an extension of the Poincaré group with spinorial generators necessary. And finally, Wess and Zumino (1974a) discovered a 4D supersymmetric field theory by trying to extend the 2D version obtained in String Theory. The results were not affected by the Coleman-Mandula result. In fact, none of the papers even referred to the Coleman-Mandula theorem, since none of them were motivated by the aim to combine internal and external symmetries.[9]

[8]See Golfand and Likhtman (1971) for the original paper and Golfand and Likhtman (1972) for an elaboration on the 1971 paper.

[9]Only in a second paper, did Wess and Zumino note in a footnote that "[t]he model described in this note, and in general the existence of supergauge invariant field theories with interaction, seems to violate $SU(6)$ no-go theorems like that proven by S. Coleman and J. Mandula [...]. Apparently supergauge transformations evade such no-go theorems because their algebra is not an ordinary Lie algebra, but has anti-commuting as well as commuting parameters. The presence of the spinor fields in the multiplet seems therefore essential" (Wess and Zumino, 1974b).

So how did they do it? An implicit assumption of the Coleman-Mandula no-go theorem is the use of Lie algebras to represent the symmetries, a mathematical assumption, which turned out to be too restrictive. Golfand and Likhtman, Akulov and Volkov as well as Wess and Zumino introduced, without explicitly realising it, a more general mathematical structure to represent symmetries, so called graded Lie algebras[10]. A structure which was introduced in the mathematics literature in the mid-1950s[11]. That more general mathematical structure allowed then to non-trivially combine internal and external symmetries in what is nowadays called supersymmetries.

4.3 The Structure of No-Go Theorems

In Sect. 4.2 we have now seen the history of a no-go theorem, from early motivations to how it was circumvented. It was chosen as a case study, as it provides us with enough detail to recognise the more general structure of no-go theorems and to identify the different elements involved. Although I will not give an argument that all no-go theorems can be put within the following abstract definition, all no-go theorems I have encountered in physics can be fit into this definition. So let us start.

The very first element of any no-go theorem is a goal G. One e.g. aims to unify internal and external symmetries, to find a hidden variable theory etc. The no-go theorem's aim is then to show that achieving this goal is not possible. Once the goal is determined the no-go theorem is set within a certain framework F. The framework can be a mathematics-framework (as in the McGlinn and O'Raifeartaigh no-go theorems), a

[10]See Chapter 2 for details on graded Lie algebras.

[11]The first paper introducing it was Nijenhuis (1955). See Corwin, Ne'eman, and Sternberg (1975) for an excellent review article on the application of graded Lie algebras in mathematics and physics.

theory-framework (like S-matrix theory), or a model-framework (based e.g. on possible extensions of existing theories like for the Bell inequalities). Within the framework one is then able to phrase the physical assumptions P that are represented by certain mathematical structures M. M for our purposes will contain both the mathematical structures used to represent the physical assumptions as well as the mathematical tools and methods used to derive the result.

In a no-go theorem it is the combination of F, P and M from which one derives something which either contradicts G directly or establishes G by violating another physical background assumption B. Taking B into account is important as we saw in O'Raifeartaigh's theorem. There one is actually able to combine internal and external symmetries but will then violate e.g. the possibility to define hermitian conjugate operators. Similarly, in the case of Bell's no-go theorem, one considers the consequences of an established hidden variable theory, the Bell inequalities, and how they disagree with the confirmed predictions of quantum mechanics. So the goal G of obtaining a hidden variable theory has been satisfied, while it disagrees with the physical background assumptions B, i.e. the predictions of quantum mechanics, which were not part of the derivation of the inequality. We have now all the components necessary to give an abstract definition:

> Definition: A *No-go result* has been established iff an inconsistency arises between
>
> - a derived consequence of a set of physical assumptions P represented by a mathematical structure M within a framework F,
> - and a goal G or a set of physical background assumptions B.

We denote an abstract no-go result with $\langle P, M, F \lightning G, B \rangle$.

The physical assumptions and the mathematical structures used to represent them are, of course, strongly dependent on each other. Obviously all elements G, B, F, P and M are dependent on each other to some extent and one may argue that it seems not obvious how to detach, for instance, P and M. But as we will see it is still reasonable to distinguish between them, since in most cases one can change the elements separately. For example I can go from a purely mathematical framework to a theory-framework while still considering the physical assumption of using spacetime symmetry and using for that purpose the mathematical structure of Lie algebras. However, as we saw in the case of the Coleman-Mandula theorem, going from one framework to the other still made it necessary to add additional assumptions to establish the no-go result. This exemplifies that one may separately change the assumptions involved; however, these changes will usually not be independent from changes in the other assumptions.

4.4 The Different Elements of a No-Go Theorem

In this section we want to discuss each element of $\langle P, M, F \lightning G, B \rangle$ in more detail. The way we have construed the abstract structure of no-go theorems they are contradictions, so as a consequence one has to deny one of its elements. In scientific practice no-go theorems have had the impact of stopping whole research programs, at least for a time. In these circumstances they were understood as showing the impossibility of G only. But given the more complex structure we have established, it is legitimate to assess the viability of denying the other elements as well. For that purpose we need to consider (i) the different elements more closely, (ii) the possible justifications we may have for each and (iii) the possible implications we may draw from their denial. We will discuss each option now in turn focusing on (i) and (ii), while turning in the next section to (iii). It is important to reiterate, although we consider each element separately, they are actually strongly dependent on each other. That is, for instance, the choice of the goal G will to a large extent inform the set of physical assumptions, while the realisation of the physical assumptions strongly depends on the mathematical structure used, and so on. We have

$$\langle P, M, F \lightning G, B \rangle \Rightarrow \neg G \vee \neg P \vee \neg B \vee \neg F \vee \neg M.$$ [12]

4.4.1 $\langle P, M, F \lightning G, B \rangle \Rightarrow \neg G$:

Here the no-go result is interpreted as the impossibility of G. This is for example how von Neumann's no-go theorem was understood for thirty years or the Coleman-Mandula theorem till the advent of supersymmetry. Although, given the general structure of no-go theorems, concentrating solely on G may seem odd, it is not surprising. G is some goal, which obviously is not yet established, while the other elements, if one is even aware of them, are at least perceived to be part and parcel of the well-confirmed physics. But if G is not part and parcel of the well-confirmed physics why is it considered to be a goal in the first place. There may be different motivations for G, which have to be analysed by considering specific examples. However, we can more abstractly recognise different possible motivations one may give.

Empirical Motivation: One motivation for setting a goal might be some

[12] This should be understood symbolically and not in a strict logical sense.

empirical observation, for which there is no adequate explanation. We saw that one motivation for combining internal and external symmetries was the problem that the mass differences between particles within one multiplet could not be accounted for. Combining internal and external symmetries was a possible way to address this.

Metaphysical Motivation: A goal may be motivated by metaphysical considerations. One way of understanding for instance the program of completing quantum mechanics, i.e. to provide a hidden variable theory, is metaphysical. Finding a theory of hidden variables is not necessitated by some observed phenomenon, that quantum mechanics cannot account for. It is motivated by the hope to find an ontologically coherent understanding of the world.

Meta-inductive Motivation: The second motivation we discussed as to why to combine internal and external symmetries was unification. Unification is also not necessitated by some empirical observations, but has been a successful ingredient in theory development. One may argue that unification is meta-inductively motivated, i.e. one infers from previous successes of attempts at unification to future ones.

Pragmatic Motivation: Another possible motivation can be purely pragmatic. Consider for instance the theorem that Nielsen and Ninomiya (1981) proved. There they show that neutrinos, or more generally chiral fermions, cannot be simulated on a lattice. So this result puts certain *calculational* limitations on simulating certain phenomenon in particle physics. As the aim of lattice gauge theories are to do certain calculations, which are otherwise very difficult, there is nothing of great foundational significance about this theorem. The original goal was pragmatically motivated.

These are possible motivations one may give for some G. There is no claim regarding the completeness of this list. The relevant point is that there may be different motivations for G and different motivations may lead to different implications one may want to draw from the no-go result. Note that there are cases where one and the same G is motivated by different theorists for different reasons, as e.g. in the example of the last section. Accordingly the implications of the same no go theorem may differ for these different theorists. For instance, it seems obvious that a goal which is metaphysically motivated may lead to a different interpretation compared to one that was motivated purely pragmatically. Laudisa (2014), for instance, argues against the significance of many recent no-go theorems in quantum mechanics. He claims that the "search for negative results [...] seems to hide the implicit tendency to avoid or postpone the really hard job", which for him is partly "to specify the ontology that quantum theory is supposed to be about" (Laudisa, 2014, p.16). However, one can understand the programme of finding a hidden variable theory, both as a metaphysical programme as well as a programme of finding a probabilistic foundation of quantum mechanics[13]. So the significance of e.g. the Bell inequalities with respect to these different goals will therefore lead to different assessments.

It is crucial to be explicit about the goal G. In two recent papers, Cuffaro (2016a) and Cuffaro (2016b), while discussing the Bell inequality and the GHZ equality, distinguishes between two kinds of context, the theoretical and the practical context. Within the theoretical context one may consider the Bell result to shed light on the questions of whether there

[13]For example, Arthur Fine (1982) followed this second route with generalised probability spaces. There were also attempts in terms of imprecise probabilities (Suppes and Zanotti, 1991; Hartmann, 2015).

is an alternative locally causal theory of the world able to replace quantum mechanics. In the practical context, on the other hand, one may ask whether one can classically reproduce, by e.g. a classical computer simulation, the predictions of quantum mechanics. These two contexts are very different. As Cuffaro shows, a denial of the goal in the theoretical context does not imply a denial of the classical simulability of the considered quantum correlations. On the other hand, as Cuffaro notes, there are quantum correlational phenomenon that no classical machine can reproduce; these then do also have no locally causal theory describing them. This illustrates an important possible general move one can make, as $\langle P, M, F \lightning G, B \rangle$ does not imply $\langle P, M, F \lightning G', B \rangle$ when $G' \vDash G$.[14]

4.4.2 $\langle P, M, F \lightning G, B \rangle \Rightarrow \neg P \vee \neg B$:

Let us turn to the physical assumptions. I include the physical background assumptions B here as well as they are after all physical assumptions. However, unlike P, if they are included at all, it is usually as a crucial assumption that is much more supported. So for that purpose we will not consider them explicitly in what follows. A no-go result that is not understood as an impossibility result of G, is quite commonly understood as an impossibility result with respect to the physical assumptions P and usually with respect to one single assumption, if one considers that one to be the least defensible assumption. This is the situation when it is argued that the Bell inequalities "prove" that the world is non-local. Physical assumptions need to be discussed case by case and a general discussion will not allow us to draw strong conclusions, but we can still recognise that there are physical assumptions of different kind.

[14] Usually, changing G will have an affect on the physical assumptions involved, as is the case in (Cuffaro, 2016b, Sect.4). However, this does not necessarily have to be the case.

Obviously, the goal G determines to a large extent the physical assumptions. If my goal is to combine the Poincaré group with some internal group, then trivially I will take as one of my physical assumptions that one of the groups is the Poincaré group. There are also independently motivated physical assumptions. When Einstein, Podolsky, and Rosen (1935) showed that quantum mechanics is incomplete, they required that the reality criterion[15] holds. There are also physical assumptions that are part of well-confirmed theories, like energy conservation, or physical assumptions that have been introduced for the sole purpose of deriving the result. An example is the analyticity assumption of Coleman and Mandula (Assumption 3).

As one can see, these different physical assumptions are not comparable in terms of the justification one can give for them. While some assumptions can be justified empirically, others cannot, but correspond to metaphysical positions and external requirements on what the future theory needs to satisfy. So while we may say that we have evidence supporting the claim that energy is conserved, we may not want to claim the same for the reality criterion in the Einstein-Podolski-Rosen setup or the factorisability assumption in the Bell inequalities. These are cases where much disagreement about the possible importance and justification for the assumption can arise and where most of the philosophical debate of no-go theorems is understandably situated. This is important, as careful analysis of these assumptions are sometimes lacking in the physics literature. For instance, Coleman and Mandula (1967, p.159) claim that the analyticity assumption "is something that most physicists believe to be a property of the real world". This, to say the least, does not seem to be

[15] As a reminder, the EPR criterion of reality states: "If, without in any way disturbing a system, we can predict with certainty (i.e., with probability equal to unity) the value of a physical quantity, then there exists an element of reality corresponding to that quantity." (Einstein, Podolsky, and Rosen, 1935, p.777)

obviously supported by empirical data.

4.4.3 $\langle P, M, F \lightning G, B\rangle \Rightarrow \neg F$:

No-go theorems are not always formulated within the framework of a theory. As we saw in the examples from the last section, the McGlinn theorem as well as the O'Raifeartaigh theorem are theorems within a mathematics-framework. That is, one considered two mathematical structures and asked whether there is a mathematical structure that non-trivially combines them. So there is no use of a theory within which this needs to be realised. On the other hand, Coleman and Mandula's theorem is a result within a theory, namely S-matrix theory. They were considering the external and internal symmetries as symmetries of the S-matrix. In other cases, one may provide a model and prove within that model-framework the no-go result.

The framework F of a no-go result has not played much of a role in the evaluation of no-go theorems. This can be due to the apparent neutrality of the framework with respect to the no-go result. In most cases it seems that the choice of framework is naturally fixed independent of the specific goal, and more by the *kind* of goal one aims to reach. In that sense it seems not to be strongly dependent on the specific goal and so seemingly neutral. If I aim to find a hidden variable account of quantum mechanics, I start by building a general model on which I impose the physical properties (elements of P) of the desired hidden variable account. That is I choose a model-framework, which may still lack the details of the dynamics of the theory etc. It is not clear how a theory-framework or a mathematics-framework could be helpful here. Similarly, it seems to be a mathematical issue, whether one can combine two symmetry groups non-trivially. So combining them without any specific theory in mind seems to be the obvious and more general approach. So one chooses a mathematics-framework. The move towards S-matrix theory, i.e. a theory-framework, was not based on not being satisfied by the mathematics-framework but was largely motivated by the aim to weaken the strong assumption of restricting oneself to finite parameter groups made by O'Raifeartaigh and it was not obvious how the theorem could have been extended to infinite parameter groups as it relied so strongly on Levi decomposition.

The above example nicely illustrates that the framework is mainly chosen for practical reasons and is not independently justified. However, using different frameworks may still provide us with different perspectives. Pitowsky (1989), for instance, provides a different perspective on the hidden variable programme. He shows that one can understand the question whether a set of probabilities are classical (Kolmogorovian) probabilities, not only by considering whether they satisfy the Kolmogorovian axioms, but also whether they satisfy a set of inequalities. He identifies for a set of probabilities the inequalities that need to hold and shows that one of them corresponds to the inequality derived by Bell. Inserting now probabilities predicted for certain[16] quantum mechanical experiments into the inequality, the inequality is violated. However, unlike in the model-framework, one infers here, within the mathematics-framework, to the comparably more mathematical conclusion that not all quantum mechanical experiments have classical probability space representations.

A reason why the significance of the framework F has not been important in the evaluation of no-go results is the lack of an obvious interpretation for $\neg F$. In the case of the goal G and the physical assumptions

[16] Not all quantum mechanical experiments lead to a violation of the Bell inequalities.

P, the denial could be understood as their respective impossibility. This is usually not so for the framework. It does not make sense to talk of the impossibility of the mathematics-framework or the model-framework, but only of the assumptions realised within it. The benefit of considering a change of framework has, however, been illustrated in the case of the hidden variable programme, where the move to a mathematics-framework presented a new perspective. The new perspective, however, came effectively with a different goal, more concerned with the probabilistic foundations rather than a locally causal hidden variable theory.

4.4.4 $\langle P, M, F \lightning G, B\rangle \Rightarrow \neg M$:

Let us turn to the last crucial element of no-go results, the mathematical structure M. M encompasses many things. There are usually many mathematical assumptions involved that are needed to obtain the result. For example, Assumption 5 of the Coleman-Mandula theorem is of this kind. It is an assumption that Coleman and Mandula admit is "both technical and ugly", and for which they hope "that more competent analysts will be able to weaken it further, and perhaps even eliminate it altogether" [p.159]. There may also be additional assumptions involved in the derivational steps, like approximations and limits. All of these can possibly be problematic and should be carefully assessed. However, we will focus on another element of M. In any representation of a problem, one uses, within a certain framework, certain mathematical structures. These are usually implicit in the derivation of the no-go result. We will concentrate on these mathematical structures for the rest of this section. More specifically, we are interested in what $\neg M$ implies. For that purpose we need to understand what the relation between the physical situation we are interested in is and the mathematical structure representing it. We are, however, not concerned with the details of the semantics of physical theories, though relevant, but take a more pragmatic attitude of the relationship between the mathematical structure and the physical situation.

Let us consider the mathematical representation of symmetries, since we are by now already acquainted with them. Symmetries are usually represented in terms of the algebraic structure of groups. There are different kinds of groups for different kinds of symmetries. The question we are interested in is whether for any single situation we are considering there is only a unique group representing it. This is not the case and has been discussed in the literature on structural underdetermination. Roberts (2011) for instance, posing it as a problem for supporters of group structural realism, shows how one can understand a group $\mathbb{G}$ as well as its automorphism group $Aut(\mathbb{G})$ as a basis from which one can construct the physical situation.[17] This, of course, goes on including the automorphism group of the automorphism group of $\mathbb{G}$ and so on. So there is a whole 'hierarchy' of symmetry groups one can consider in representing the physical situation.

This, as we by now should know, is not the only option. Note, both $\mathbb{G}$ and $Aut(\mathbb{G})$, although different groups, are still the same algebraic structure, in the sense that they both satisfy the same algebraic axioms, namely those of groups. There are, however, many algebraic structures we could in principle use to represent the same physical situation. As we saw in the case of supersymmetry, it was exactly this move from one algebraic structure, namely Lie algebras, to another algebraic structure, namely $\mathbb{Z}_2$-graded Lie algebras, that allowed internal and external symmetries to combine non-trivially. Similarly, we require from probabilities that they

[17] This is not true for all groups as some groups, e.g. the permutation group S_3, is isomorphic to its automorphism group. See Roberts (2011, p.62) for more details.

satisfy the Kolmogorov axioms. As we saw in the previous section, certain quantum mechanical probabilities violate the axioms of Kolmogorov. We do not want to say that they are therefore not probabilities but instead that they may satisfy different axioms, i.e. they are different probabilistic structures.

Another possible move in mathematical structure is possibly realised within the theory-framework, namely when moving from one formulation of the theory to another. We can consider a no-go result we obtain in one formulation to also hold in the other, only if we have reason to believe that they are equivalent in the relevant sense. However, the Coleman-Mandula result is a result within S-matrix theory, and it is, for example, not obvious that it will similarly hold within Lagrangian quantum field theory, as the symmetries of the S-matrix are not necessarily symmetries of the Lagrangian. The difference in mathematical structure is something much discussed in the context of different formulations of classical mechanics. There one is concerned with the question whether the Lagrangian or the Hamiltonian formulation is the 'right' structure. Jill North (2009) has argued that Hamiltonian mechanics ascribes 'less' structure to the world than Lagrangian mechanics does and from that she "infer[s] that Hamiltonian mechanics is more fundamental than Lagrangian mechanics"[p.76] based on this simplicity comparison. This has subsequently been criticised by Curiel (2014, p.303), who argues that it is not clear what notion of simplicity has been used by North and even if it would be made precise there is not a unique definition. Different definitions will lead to different simplicity assignments. Barrett (2015, p.816), similarly, gives a definition of how to compare sizes of structures and argues that the structures of Lagrangian and Hamiltonian mechanics are actually incomparable.

This should be clear even intuitively. We mentioned in Sect.2.3.2 that graded Lie algebras can be understood as generalisations of Lie algebras. In this sense, everything a Lie algebra can describe can also be described by a graded Lie algebra; the converse however is not true. Here we have a case where we see how one structure is richer than the other. Let us consider this in the context of the Kolmogorov axioms. If we want to change the structure we can consider weakening one of the axioms, e.g. the additivity axiom, leading to what is sometimes called upper or lower probabilities. This will similarly count as a more generalised structure in the sense that all Kolmogorovian probabilities will satisfy these changed axioms as well. Another option, however, would have been to not require probabilities to be necessarily positive. This again would still allow us to account for all Kolmogorovian probabilities. However, how would one compare negative probabilities with upper or lower probabilities? It is at least plausible that there will not be a unique natural measure of comparison. This example also illustrates how one can, in principle, always come up with other structures, which are not intuitively comparable. There are two points I should make at this point. First, I am not saying that intuitively incomparable structures cannot be compared in a precise way. I am only saying, following Curiel, that the precise choice of comparison will depend on assumptions about simplicity and parsimony, which are not empirically justifiable. Second, changing algebraic structures, in the above sense, does not imply that all these different structures are in some sense physically unproblematic. It might very well be that the structures will violate other fundamental principles and are therefore not adequate to replace the previous structure. However, this is not a result which is entailed by the no-go theorem itself and requires independent analysis.

4.5 What are No-Go Theorems Good for?

No-go theorems are complicated beasts, which as we have seen are hard to dissect. We have provided a possible abstract definition of no-go theorems, which allowed us to analyse it in more detail. We would now like to draw some more general conclusion, which we establish through several claims based on arguments we have given before.

Claim 1: No go theorems are more complicated than usually assumed. As we have seen, they are usually posed as either impossibility results with respect to the goal G or some element of the physical assumptions P. This simplified picture ignores the crucial role played by the framework F and the mathematical structure M.

Claim 2: No-Go theorems do not come equipped with an ordinal preference assignment on its elements. A no-go theorem is a contradiction, which derives from a set of elements. The result itself does not say, which of the elements involved in the derivation is more and which one is less justified.

Claim 3: What No-Go theorems imply depends on one's ordinal preference assignment. Once we have established a no-go theorem we need to address the question, how we wish to address the contradiction, i.e. how we wish to interpret the no-go result. The interpretation depends on which of the elements of the no-go theorem we are most willing to change or give up on. However, as we have seen, not all elements are empirical certainties of nature, but vary strongly based on the justifications one can give for them. Furthermore, different scientists may have different justifications for the elements of a no-go theorem, corresponding to a difference in ordering of what one prefers to give up or change first. This difference in preference assignment will correspond to differences in interpreting the same no-go theorem. There is not a unique implication one can draw from the no-go theorem by itself.

Claim 4: The mathematical structure M is usually the least justifiable of the elements. In principle, we can imagine an empirically motivated goal G and similarly empirically well-confirmed physical assumptions P, which determine the framework F. We cannot claim the same for mathematical structures. While one may be committed to a certain goal and physical assumptions, this is usually less so with the mathematical structures. We may have many good reasons to choose one mathematical structure rather than another, based on simplicity and naturalness assumptions. But the empirical access to them is very limited. Keeping certain physical assumptions fixed one can empirically only point to the insufficiency of a certain mathematical structure to account for some observed phenomenon. This leaves a whole lot of weaker and therefore more encompassing structures untouched. The space of all mathematical structures is not a clearly defined space. As such, it does not allow for a rigorous "working though all structures"-approach, but only allows for theoretical exploration. This naturally leads to the methodological implications of no-go theorems, which come in our next claim.

Claim 5: No-go theorems are best understood as go theorems. No-go theorems usually do not strictly speaking allow for an interpretation as an impossibility result with respect to G or P, as that would imply one has certainty with respect to the rest of the elements and this is, as we saw above, not the case. So what do they imply? If we, for instance, accept the mathematical structure M as the weakest element, we interpret the no-go theorem as implying $\neg M$. But as we have already said, $\neg M$ cannot meaningfully be interpreted as the impossibility of the mathematical structure, but as an invitation to consider alternative mathematical structures to replace it. This, of course, does not imply that the alternative mathematical structures do not lead to no-go theorems themselves. It also does not imply, that one should never end the search for alternative mathematical structures. It only implies that it is the first line of attack in interpreting the no-go result. This can and usually will come with an adjustment and an exploration within alternative P's. It is in this sense that no-go theorems are one of the best tools for achieving one's goal G as they outline possible approaches to reach it (not only via $\neg M$). They are excellent tools in theory development, while being unreliable tools in stopping research programmes.

4.6 Conclusion

We started with a case study of the development of a no-go theorem from particle physics, which provided us with enough detail to recognise the different abstract elements of no-go theorems. We discussed each element in detail coming to the conclusion that no-go theorems cannot be understood as impossibility results in the strict sense. Especially, the mathematical structure M poses a threat to this strong conclusion. This turned the role of no-go theorems around. Rather than understanding no-go theorems as providing us deep insights about what is and what is not possible in the world, they should be understood as a methodological starting point in theory development.

While we have outlined a more systematic analysis of no-go theorems, we could have chosen an alternative route, namely, via meta-induction on the history of physics. Von-Neumann's No-go theorem was superseded by both, actual hidden variable theories (pilot wave theories, Bohmian mechanics), and further no-go theorems where both the physical assumptions P as well as the mathematical structures M have been changed. The impossibility to simulate chiral fermions on a lattice, the Nielsen-Ninomiya theorem, was circumvented via the introduction of domain wall fermions by extending the mathematical representation of the lattice with an additional dimension (Kaplan, 1992; Shamir, 1993). Weinberg and Witten (1980) proved that e.g. gravitons cannot be composite particles in a relativistic quantum field theory. There is now a whole plethora of counter examples from conformal field theories, massive gravity to String theory[18]. We have already discussed Supersymmetry and how it circumvented the Coleman-Mandula theorem. We could continue with other examples, but this should suffice for our purposes. One could now argue, based on this historical evidence, that maybe current no-go theorems will be superseded by ways to circumvent them as well. This is in complete agreement with our analysis above. That is, it was to be expected that no-go theorems do not say the last word with respect to one's goal G. Our analysis actually provides the explanation why they do not. However, history is also full of examples where these no-go theorems did actually have the effect of stopping whole research programmes. That is, we have many historical examples where no-go theorems were systematically misunderstood in what they can imply. So no-go theorems have played a role in the history and methodology of physics, for which they did not provide the argumentative support. There is a discrepancy between what no-go theorems *can* imply and how they were actually interpreted in practice. Recognising what they can imply provides us with a more adequate use of them as a tool in theory development. This more adequate use is the understanding that no-go's are actually the best go's!

[18]See (Bekaert, Boulanger, and Sundell, 2012) for a review article on how the Weinberg-Witten theorem is circumvented in these theories.

Appendix: Short Reminder on the Role of Symmetries in Particle Physics

One reason why symmetries are considered to be at the least a powerful heuristic tool is their ability to give (i) (in some sense) rise to the characterisation of the particle content and (ii) the dynamics of the theory. I will now briefly review the first part and what is usually meant by that, and ignore (ii) since it is not relevant for our discussion[19].

The oft-cited statement by Ne'eman and Sternberg nicely illustrates the importance physicists associate with the relation between groups and particle characterisation:

> "Ever since the fundamental paper of Wigner on the irreducible representations of the Poincaré group, it has been a (perhaps implicit) definition in physics that an elementary particle 'is' an irreducible representation of the group, G, of 'symmetries of nature'." (Ne'eman and Sternberg, 1991, p.2)

Let us briefly discuss how particle properties and symmetry groups are related which is at the basis of the above statement.[20] Let us start with a general account before considering the example of the Poincaré symmetry that we will need later on.

1. Start by specifying a symmetry group G
2. Consider unitary representations $U(g)$ with $g \in G$ infinitesimally, i.e. $U = 1 + i\epsilon^i T^i + ...$

[19] See e.g. Teller, 2000,Healey, 2005, Afriat, 2013 for some discussion on this topic.
[20] This is discussed in most standard texts on quantum field theory, especially clearly in (Weinberg, 1996).

3. Calculate the algebra associated with the generators T^i, i.e. $[T^i, T^j] = i \cdot f^{ijk} T^k$.

4. Find the Casimir operators C^α which satisfy $[C^\alpha, T^i] = 0$ for all i.

5. The eigenvalues of the Casimir operators are the invariant properties of the particles (uniquely determining the irreducible representations of a group).

In the concrete case of the Poincaré group $\mathcal{L}$ we have group elements $g = (\Lambda, a)$, corresponding to Lorentz transformations and translations respectively. Infinitesimally these correspond to the transformations:

$$\begin{aligned} \Lambda^\mu_\nu &= \delta^\mu_\nu + \omega^\mu_\nu + \ldots \\ a^\mu &= \epsilon^\mu + \ldots \quad . \end{aligned}$$

This gives rise to the unitary representations $U(1 + \omega, \epsilon) = 1 - \frac{i}{2}\omega_{\mu\nu} M^{\mu\nu} + i\epsilon_\mu P^\mu$ with the Lorentz generators $M^{\mu\nu}$ and translation generator P^μ for which the algebra has to be determined[21]. It turns out that there are two operators commuting with all the generators of the algebra yielding the following Casimirs P^2 and W^2 with $W_\mu = \frac{1}{2}\epsilon_{\mu\nu\rho\sigma} M^{\nu\rho} P^\sigma$ being the Pauli-Lubanski pseudovector. The eigenvalues of these two Casimir operators, and thereby the characterising features of the particles, are mass m and spin s. However, as we know, these characterising properties of the particles are not sufficient to characterise the properties of particles like quarks. This leads to further internal symmetry groups leading to properties like the different quark flavours and so on. The steps are completely analogous to the Poincaré case.

[21] The specific form of the algebra should not concern us here.

Chapter 5

Confirmation via Analogue Simulation

5.1 Introduction

In this chapter we want to turn our attention away from non-empirical theory assessment and concern ourselves again with experiments. However, with experiments on systems, which do not fall under the domain of the theory we aim to test.[1] We will articulate a refinement and extension of existent analysis of the role of analogies in science (Keynes, 1921; Hesse, 1964; Hesse, 1966; Bartha, 2010; Bartha, 2013) inspired by fluid mechanical 'dumb hole' analogues to gravitational black holes (Unruh, 1981; Novello, Visser, and Volovik, 2002; Barceló, Liberati, and Visser, 2005; Unruh, 2008). Our central claim, which we take to be both bold and well founded, is that this case exemplifies a notion of *analogue simulation* that, unlike other species of *analogical reasoning*, has the potential to provide a conduit for confirmation. Trading on a robust *syntactic isomorphism* between the relevant modelling frameworks, analogue simulation allows certain inaccessible phenomenology in the target system to be probed by

[1]This chapter corresponds to (Dardashti, Thébault, and Winsberg, forth.), which has been accepted for publication by the British Journal for the Philosophy of Science.

experimentation on the analogue. Given further model-external and empirically grounded arguments, this then allows us to 'confirm' the existence of novel phenomenology in the target system via the observation of its correlate in the analogue. The potential importance of this claim is particularly startling in the context of our chosen example since Hawking radiation is among the gravitational phenomena that 'dumb holes' have the capacity to simulate, and by our lights confirm. Thus, if our analysis is correct, the quantum phenomenology of black holes is potentially within reach of contemporary experimental research in analogue gravity (Carusotto et al., 2008).

Our arguments regarding confirmation via analogue simulation do, however, cut both ways. Given the requirement for additional model-external and empirically grounded arguments, on its own the analogue simulation of some phenomena is not taken to be confirmatory. Thus, claims made in the literature (Weinfurtner et al., 2013) regarding the 'verification' of classical aspects of Hawking radiation must, for the moment, be treated carefully. On our analysis, confirmation of Hawking radiation via analogue simulation can only be established given the acceptance of a chain of reasoning involving universality arguments in combination with diverse realisations of the counterpart effect. These stringent conditions provide both the normative thrust of our analysis of the case in hand and a framework for the investigation of further cases. In general terms, our aim in what follows is to establish the existence of a distinct notion of analogue simulation, and then to provide conditions for this mode of scientific inference to be confirmatory. Whether or not there exist cases in which these conditions in fact obtain remains to be seen: further progress in empirical science is needed before one could justifiably claim confirmation of Hawking radiation via analogue simulation.

In the following Section we will briefly review the physical background necessary for a basic understanding of: Hawking radiation in semi-classical gravity (Sect. 5.2.1); the modelling of sound in fluids (Sect. 5.2.2); and the acoustic analogue model of Hawking radiation (Sect. 5.2.3). Section 3 then contains explication of analogue simulation and our claim that it can provide a means for confirmation. In the first subsection (Sect. 5.3.1) we review the traditional notion of analogical reasoning, introduce a framework for understanding analogue simulation, and then contrast the two. In the following subsection (Sect. 5.3.2) we deal with the problem of justifying the inferences necessary for analogue simulation to enable confirmation. Our key idea is that in certain circumstances predictions concerning inaccessible phenomena can be confirmed via an analogue simulation in a different system. As we shall see, one is only justified in making such claims once one has established additional empirically grounded and model-external arguments for the accuracy and robustness of the relevant modelling frameworks and syntactic isomorphism within the domains involved. We conclude Section 3 with a recapitulation of the key distinctions and their relevance to issues of confirmation (Sect. 5.3.3). The problems of experimental realisation of Hawking radiation (Sect. 5.4.1) and of finding the relevant model-external, empirically grounded arguments for the dumb hole/black hole case (Sect. 5.4.2) then become our main occupation. In completion of our argument, we present the case for the dumb hole/black hole correspondence offering us the possibility for confirmation of Hawking radiation via analogue simulation (Sect. 5.4.3). We conclude by offering a prospectus for extension of the idea of analogue simulation to other areas of science, and give a short sketch of one case of particularly obvious relevance (Sect. 5.5).

5.2 Physical Background

5.2.1 Hawking Radiation in Semiclassical Gravity

In this section we will give a brief overview of the basis behind Hawking's famous calculation demonstrating that black holes are associated with a particle flux connected with a characteristic 'black hole temperature' (Hawking, 1975). Before we do this we should note that a number of rich and important interpretational issues regarding such *Hawking radiation* are yet to be fully resolved. Most troublingly, the sense in which the radiation is localisable to the interior of the black hole, the event horizon, or the exterior of the black hole, is far from clear. Hawking's calculation *does not* provide a causal mechanism for the radiation and thus, to a degree, renders the phenomena rather mysterious. This notwithstanding, the original thermal model for Hawking radiation has proved 'remarkably robust' under the inclusion of various complicating factors (Leonhardt and Philbin, 2008; Thompson and Ford, 2008), and thus can be taken as sufficient for our purposes.

Our starting point is a semi-classical approach to gravity where we consider a quantum field within a fixed spacetime background. We will follow the standard treatment[2] and consider the simplest possible model with a scalar field ϕ considered in 1+1 Minkowskian background. The classical wave equation for a free scalar field is simply given by $g^{ab}\nabla_a\nabla_b\phi = 0$ where the scalar field is a quantum operator, i.e., obeys the canonical equal-time commutation relations and acts on a suitably constructed Hilbert space. We then expand the scalar field in a basis of

[2] See (Mukhanov and Winitzki, 2007) for an elementary introduction.

orthonormal plane wave solutions:

$$\phi = \int d\omega (a_\omega f_\omega + a_\omega^\dagger f_\omega^*) \tag{5.1}$$

where $f_\omega = \frac{1}{\sqrt{2}} e^{-i(\omega t - kx)}$ and a_ω, $a_\omega^\dagger$ are operators now satisfying $[a_{\omega'}, a_\omega^\dagger] = \delta(\omega' - \omega)$, for some frequency ω and real constant k. The operators $a_\omega^\dagger$ and a_ω can thus be interpreted as creation and annihilation operators. We can now consider the vacuum state for the scalar field at past null infinity, $\mathcal{J}^-$, to be the 'in' state, and define it in the usual way as $a_\omega |0\rangle_{in} = 0$ for all $\omega > 0$. It is natural to then also define the number operator for the 'in' state at each frequency ω as $N_\omega^{in} = a_\omega^\dagger a_\omega$.

Now, consider an alternative set of solutions $\{p_\omega, p_\omega^*\}$ which form a complete orthonormal basis and with respect to which we can also expand the scalar field:

$$\phi = \int d\omega (b_\omega p_\omega + b_\omega^\dagger p_\omega^*) \tag{5.2}$$

with $[b_{\omega'}, b_\omega^\dagger] = \delta(\omega' - \omega)$. We use these creation and annihilation operators defined in this new basis to specify the properties of the 'out' state at future null infinity, $\mathcal{J}^+$, via $b_\omega |0\rangle_{out} = 0$, and $N_\omega^{out} = b_\omega^\dagger b_\omega$. Since the massless scalar field is completely determined by its Cauchy data on either of the surfaces $\mathcal{J}^-$ or $\mathcal{J}^+$, it can be expressed in the form (1) or (2) everywhere. This means that we can transform the f_ω solutions in terms of the p_ω solutions, and vice versa, via the *Bogoliubov transformations*:

$$f_\omega = \int d\omega' (\alpha_{\omega\omega'}^* p_{\omega'} - \beta_{\omega\omega'} p_{\omega'}^*) \qquad p_\omega = \int d\omega' (\alpha_{\omega\omega'} f_{\omega'} + \beta_{\omega\omega'} f_{\omega'}^*) \tag{5.3}$$

where $\alpha_{\omega\omega'}$ and $\beta_{\omega\omega'}$ are the Bogoliubov coefficients given by the inner products $\alpha_{\omega\omega'} = (p_\omega, f_{\omega'})$ and $\beta_{\omega\omega'} = -(p_\omega, f_{\omega'}^*)$. Similarly, the Bogoliubov coefficients relate the relevant mode operators to each other, e.g.

$$b_\omega = \int d\omega' (\alpha^*_{\omega\omega'} a_{\omega'} - \beta^*_{\omega\omega'} a^\dagger_{\omega'}). \tag{5.4}$$

These results give us the basis to calculate the expectation value of the *out* number operator for the *in* vacuum state:

$$_{in}\langle 0|(N^{out}_\omega)|0\rangle_{in} = \quad _{in}\langle 0|b^\dagger_\omega b_\omega|0\rangle_{in} = \int d\omega' |\beta_{\omega\omega'}|^2 \tag{5.5}$$

Thus, even in this simple model for semi-classical gravity we can see that the initial vacuum state of a scalar field in a classical spacetime need not appear as a vacuum state to observers at positive null infinity: it may contain a flux of 'out-particles' which one can calculate simply by determining the relevant coefficient $\beta_{\omega\omega'}$. In the case where the two solutions are related to each other by a Lorentz transformation, both observers associated with the solutions will of course agree on the number of particles observed and (5.5) will vanish.

What Hawking's calculation (Hawking, 1975, pp. 204-213) shows is that, for a spacetime which features the establishment of an event horizon via gravitational collapse leading to a black hole, one can derive the asymptotic form of the relevant coefficients $\beta_{\omega\omega'}$, and show that it depends only upon the *surface gravity* of the black hole and in the long-time limit not on the details of the gravitational collapse.[3] We can define surface gravity, κ, in terms of the magnitude of the acceleration, with respect to Killing time, of a stationary zero angular momentum particle just outside the horizon (Jacobson, 1996). This is, more intuitively put, the force

[3]In essence the calculation follows the same lines as that for the scalar field ϕ in 1+1 Minkowskian space, only the fields are also expanded for the surface defined by the event horizon. Importantly one does not need to consider a quantum field in a curved spacetime since the field operators are only evaluated in the asymptotic regime which is presumed to be flat.

per unit mass that must be applied at infinity in order to hold the particle on its path. For a nonrotating neutral black hole the surface gravity is given by $\frac{1}{4M}$, where M is the black hole mass. The role of surface gravity in black hole thermodynamics is almost identical to that of temperature in conventional thermodynamics. In fact, we define the black hole temperature, T_{BH}, in terms of the surface gravity. Hawking's calculation thus gives a general demonstration that there is a connection between the intrinsic thermodynamic properties of a (non-eternal) black hole (or at least its horizon) and a non-zero particle flux at late times. The precise relation takes the form:

$$\langle N_{\omega}^{\text{Black Hole}} \rangle = \frac{1}{e^{\frac{2\pi\omega}{\hbar\kappa}} - 1} \qquad T_{BH} = \hbar\kappa/2\pi \tag{5.6}$$

where we now simply refer to the evaluation of the expectation value of the late time particle number operator relative to the initial vacuum as the 'black hole' particle flux. We thus see that the basis behind Hawking's derivation of Hawking radiation is a very general one: it requires us only to consider simple features of quantum vacuum states when evaluated in classical spacetime backgrounds which feature Killing horizons. It is a kinematical effect not a dynamical one. The Einstein equation is not used anywhere in the calculation. As noted above, such a derivation of what is, after all, a highly non-trivial *physical* effect seems a little unsatisfactory from a causal view point and it has some severe problems some of which we discuss later[4]. However, for our purpose of investigating analogue models of Hawking radiation it will prove best to think about the gravitational derivation in precisely such general and non-microcausal terms.

[4]For a recent discussion of some of these problems see (Unruh, 2014).

5.2.2 Modelling Sound in Fluids

Sound is generally understood as a small vibratory or wavelike disturbance in a medium. Classical physics deals with such acoustic phenomena in both fluids and solids and, within certain realms of application, gives an empirically adequate description. A simple but powerful classical acoustic model (Landau and Lifshitz, 1987, §64) is that where the fluid is taken as a continuous, compressible, inviscid medium and sound is understood as a longitudinal oscillatory motion with small amplitude within the medium. Since the fluid in such a model is treated as a continuous medium, 'points' within it are really volume elements that are presumed to contain a very large number of fluid molecules. The volume elements are taken to be *very small* with respect to the overall fluid volume, and *very large* with respect to the inter-molecular distances, meaning that the model is only valid in a particular window of applicablity determined by the size of the fluid and relevant molecular distances. The first fundamental equation within this model of a fluid is the *continuity equation*. This is simply an expression of the conservation of matter, and takes the form:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \vec{v}) = 0 \tag{5.7}$$

where ρ is the mass density of the fluid at a particular point, and $\vec{v}$ is the velocity of the fluid volume element. A third quantity, in addition to density and velocity, that we use to characterise the fluid is the pressure, p. The total force due to the surrounding fluid acting on a unit of fluid – i.e. a fluid 'particle' – is given by $-\nabla p$ and this, by Newton's second law, must be equal to the rate of change in velocity of the fluid particle relative to space. These considerations lead us to the second fundamental equation, the Euler equation:

$$\rho\Big(\frac{\partial \vec{v}}{\partial t} + (\vec{v}\cdot\nabla)\vec{v}\Big) = -\nabla p \tag{5.8}$$

The simultaneous solution of these two equations gives us a basic model for the entire fluid flow, which is characterised in a given situation by the triple of functions $\rho(x, y, z, t)$, $p(x, y, z, t)$ and $\vec{v}(x, y, z, t)$.

We understand a sound wave in terms of an alternate compression and rarefaction at each point in the fluid, and can produce a model for sound traveling through a fluid in terms of the movement of small fluctuations around the equilibrium density and pressure. Explicitly we can consider *linearized* fluctuations around the exact solutions of the form:

$$\begin{aligned}
\rho(t,x,y,z) &= \rho_0(t,x,y,z) + \epsilon\rho_1(t,x,y,z) + ... \\
p(t,x,y,z) &= p_0(t,x,y,z) + \epsilon p_1(t,x,y,z) + ... \\
\vec{v}(t,x,y,z) &= \vec{v}_0(t,x,y,z) + \epsilon\vec{v}_1(t,x,y,z) + ...
\end{aligned}$$

The equations of motion of the fluctuations described by $(\rho_1, p_1, \vec{v}_1)$ is then precisely the model for sound propagation in fluids. Such a model (given additional simplifying assumptions) allows us to reproduce a number of important fluid acoustic phenomena such as resonance and reflection/refraction of sound waves in mediums composed of homogenous layers of different fluids (Landau and Lifshitz, 1987, §65-81) .

As emphasised already, the model's applicability depends crucially upon the fluid volume, volume unit size, and inter-molecular distances all being orders of magnitude apart. Further limits on the applicability are also given by the speed at which both the sound wave and the fluid itself are travelling: if either of these are comparable to the speed of light then a relativistic fluid dynamical model would be needed (Landau and Lifshitz, 1987, §134). Now, the point crucial to our analysis is that even if we assume that *in these respects* the model *is* within its domain of applicability, there is still scope for it to need modification due to quantum effects. If we consider a Bose-Einstein condensate (realised, for example in terms of a superfluid such as liquid He^4 close to absolute zero) then certain assumptions of our simple model for sound propagation in fluids necessarily breakdown, but certain do not. In particular, within certain regimes, the quantum field corresponding to the Bose-Einstein condensate can be separated into a bulk flow component and linearized fluctuations. We can then still treat the bulk as an essentially classical, macroscopic fluid similar to those discussed above,[5] but we now treat the linearized fluctuations (i.e. sound) as fundamentally quantum mechanical. That is, under certain conditions, for certain very low temperature fluids, even if the bulk fluid flow is still treated classically, it becomes appropriate to treat the elementary excitations as quanta of the sound wave – i.e. as phonons akin to those used in the quantum mechanical descriptions of sound in crystalline solids. This is a strikingly similar type of semi-classical approximation to that made by Hawking in the context of black holes as discussed above. In the next section we will see that the similarity between acoustic and gravitational models can, in certain circumstances, be arranged so that they have *exactly the same* mathematical structure. This will lead the way to the introduction of the notion of *Analogue Simulation* in Section 3.

[5] In fact the bulk fluid equations can be shown to reduce to a continuity equation plus an Euler equation which are completely equivalent to those of a classical inviscid fluid apart from the existence of the quantum potential term in the latter (Barceló, Liberati, and Visser, 2005, p.60).

5.2.3 The Acoustic Analogue Model of Hawking Radiation

Following (Unruh, 1981) and (Barceló, Liberati, and Visser, 2005), let us consider a simple model for a classical fluid along the lines discussed above (i.e., a continuous, compressible, inviscid medium in a non-relativistic regime), but with the additional assumptions that the fluid is *barotropic* and *locally irrotational*. Barotropic simply means the pressure is a function of the density (or vice versa), and so equates to imposing the condition $p = p(\rho)$. Locally irrotational means that their are no vortices in the fluid and equates to imposing the condition $\nabla \times \vec{v} = 0$, which implies $\vec{v} = \nabla\psi$, where we have introduced the velocity potential ψ. The barotropic and irrotational conditions allow us to much simplify Euler's equation so that it reduces to a form of the Bernoulli equation. We can then consider the linearisation of the solutions to this equation of motion for the entire fluid about a background, (ρ_0, p_0, ψ_0),

$$\begin{aligned} \rho(t,x) &= \rho_0(t,x) + \epsilon\rho_1(t,x) + ... \\ p(t,x) &= p_0(t,x) + \epsilon p_1(t,x) + ... \\ \psi(t,x) &= \psi_0(t,x) + \epsilon\psi_1(t,x) + ... \end{aligned}$$

Again we identify the sound waves in the fluids with the fluctuations (ρ_1, p_1, ψ_1) about the background, which is interpreted as bulk fluid motion. The linearised version of the continuity equation (together with the barotropic condition) then allows us to write the equation of motion for the fluctuations as :

$$\frac{\partial}{\partial t}\left(\frac{\rho_0}{c^2_{\text{sound}}}(\frac{\partial\psi_1}{\partial t} + \vec{v_0}\cdot\nabla\psi_1)\right) = \nabla\cdot\left(\rho_0\nabla\psi_1 - \frac{\rho_0\vec{v_0}}{c^2_{\text{sound}}}(\frac{\partial\psi_1}{\partial t} + \vec{v_0}\cdot\nabla\psi_1)\right) \tag{5.9}$$

This equation can be rewritten as:

$$\frac{1}{\sqrt{-g}}\frac{\partial}{\partial x_\mu}(\sqrt{-g}g^{\mu\nu}\frac{\partial}{\partial x^\nu}\psi_1) = 0, \tag{5.10}$$

where we have defined the acoustic metric

$$g_{\mu\nu}^{\text{acoustic}} = \frac{\rho_0}{c_{\text{sound}}}\begin{pmatrix} -(c_{sound}^2 - v_0^2) & \vdots & -(v_0)_j \\ \dots & \cdot & \dots \\ -(v_0)^i & \vdots & \delta_{ij} \end{pmatrix} \tag{5.11}$$

Note that (5.10) reduces to the simple free scalar field equation we considered in Section 5.2.1 for a $1+1$-Minkowski metric. We thus see that the propagation of sound in a fluid can be understood as being governed by an acoustic metric of the form $g_{\mu\nu}$. The acoustic perturbations couple only to the effective acoustic metric and not to the physical spacetime metric which describes the spacetime in which the fluid exists. Formally the acoustic metric describes a (3+1)-dimensional Lorentzian (pseudo-Riemannian) geometry and it depends algebraically on the density, velocity of flow, and local speed of sound in the fluid meaning it is constrained to have at most three degrees of freedom per point in spacetime.

The close similarity between the acoustic case and gravity can be seen immediately if we consider the Schwarzschild metric in Painleve-Gullstrand coordinates (within which the Schwarzschild geometry is written such a way that space is flat, even though spacetime is curved). This takes the form:

$$g_{\mu\nu}^{\text{Schwarzschild}} = \begin{pmatrix} -(c_0^2 - \frac{2GM}{r}) & \vdots & -\sqrt{\frac{2GM}{r}}\vec{r_j} \\ \dots & \cdot & \dots \\ -\sqrt{\frac{2GM}{r}}\vec{r_i} & \vdots & \delta_{ij} \end{pmatrix} \tag{5.12}$$

This similarity can be transformed into an isomorphism (up to a factor) under certain very specific conditions. Explicitly, following (Novello, Visser, and Volovik, 2002), we make the restrictions that: i) c_{sound} is a position-independent constant; ii) the fluid moves radially with a velocity profile $v_{\text{fluid}} = \frac{1}{\sqrt{r}}$; and iii) the background density ρ_0 is a position-independent constant. Given these requirements our fluid metric becomes such that it is proportional to the Schwarzschild metric above. The role of the black hole event horizon is now played by the effective acoustic horizon where the inward flowing magnitude of the radial velocity of the fluid exceeds the speed of sound. The black hole is replaced by a *dumb hole*.

Thus we have a methodology for *simulating* (in a sense to be discussed more fully shortly) a classical black hole using fluid mechanics. This is in of itself a rather impressive result. However, one is able to stretch the fluid/gravity analogy even further and consider *both* classical and quantum mechanical acoustic phenomena within the fluid as analogies to radiative phenomena within a black hole spacetime. The relevant calculation for the acoustic case proceeds in precisely the same manner as the Hawking calculation considered above only with the scalar field corresponding to sound, and in this case leads to a relation between the late time sound flux associated with the dumb hole, and the 'surface gravity' of the acoustic horizon. The latter is simply equal to the physical acceleration of the fluid as it crosses the event horizon, a_{fluid}, and is given by an expression of the form:

$$\kappa = c_{\text{sound}} \left| \frac{\partial v}{\partial n} \right| = a_{\text{fluid}} \tag{5.13}$$

Where $\frac{\partial}{\partial n}$ is the normal derivative of the fluid velocity as it crosses the event horizon. Given this, we can make numerical estimates for the acoustic Hawking Temperature, T_H^{acoustic}, of a dumb hole for any given fluid. And since, unlike for black holes, we have experimental access to fluids, it means that there is in principle a means for testing the predicted Hawking radiation of a dumb hole.

We will return to the various interesting issues surrounding experimental observation of Hawking radiation via analogue models in Section 4. There we will consider the extant experiments to test for fluid mechanical Hawking radiation, and also give some details regarding possible future experiments using different analogical models. We will also consider the implications of the short length scales breakdown of the analogue models for our understanding of the possible Planckian breakdown of the gravitation model. This will finally lead into a discussion of Hawking radiation in the context of *universality*, and allow us to understand precisely what dumb holes could tell us about gravity. Before then, in the following section, we will consider the implications of the analogue models we have been discussing for the general analysis of *simulation* and *analogy* in physical theory.

5.3 Simulation and Analogy in Physical Theory

5.3.1 Analogical Reasoning and Analogue Simulation

Arguments by analogy are very common in both science and philosophy. An oft-cited example is an argument offered by the Scottish Philosopher Thomas Reid for the existence of life on other planets:

> Thus, we may observe a very great similitude between this earth which we inhabit, and the other planets, Saturn, Jupiter, Mars, Venus, and Mercury. They all revolve round the sun, as the earth does, although at different distances, and in different

> periods. They borrow all their light from the sun, as the earth does. Several of them are known to revolve round their axes like the earth, and, by that means, must have a like succession of day and night. Some of them have moons, that serve to give them light in the absence of the sun, as our moon does to us. They are all, in their motions, subject to the same law of gravitation as the earth is. From all this similitude, it is not unreasonable to think, that those planets may, like our earth, be the habitation of various orders of living creatures. There is some probability in this conclusion from analogy (Reid and Hamilton, 1850, pp. 16-17)

In light of this quotation, let us pose a simple question: are the inferences one can make about black hole Hawking radiation by drawing on observations of dumb holes of the same kind as the inferences one can make about the existence of life on other planets by drawing on observations of life on Earth? We think not, and to make this clear we would like to draw a distinction between a notion of *analogical reasoning*, on the one hand, and a second notion of *analogue simulation*, on the other.

The literature on analogical reasoning is fairly extensive, with particularly noteworthy contributions by Keynes (Keynes, 1921), Hesse (Hesse, 1964; Hesse, 1966) and Bartha (Bartha, 2010; Bartha, 2013). Drawing on this literature (in particular (Bartha, 2013)), we can characterize analogical inferences in the following way: First, call the less accessible system about which we hope to make inferences the target, T, and call the more accessible system we hope to make use of in analogy the source, S. An argument using analogical reasoning then takes the following form (Bartha, 2013):

P1. S is similar to T in certain *known* respects.

P2. S has some further feature Q.

C. Therefore, T also has the feature Q, or some feature $Q^{\star}$ that is similar to Q.

This seems to fit the Reid case rather well.

Following Hempel (Hempel, 1965), the literature also recognizes a particular kind of argument by analogy, what Hempel called *nomic isomorphism*. Hempel characterized this kind of reasoning as a situation in which S and T are each governed by a set of laws, between which there is a *syntactic isomorphism*. When talking about syntactic isomorphism '[t]he essential idea is that the two sets of physical laws have a common mathematical form and may be obtained by assigning different physical interpretations to the symbols that appear in that common form' (Bartha, 2010, pp. 208-9).[6]

These ideas offer a solid starting point for the philosophical analysis of the dumb hole case; however, several amendments and clarifications are required.

One central recommendation is that Hempel's notion of nomic isomorphism, which is a sub-species of analogical reasoning, should be replaced with a broader concept of *analogue simulation*. Analogue simulation, as we understand it, can occur even when the syntactic isomorphism one can identify does not hold between the laws governing the two systems in generality. This can be seen clearly in the dumb hole case

[6]We retain Hempel's 'syntactic isomorphism' principally for historical reasons even though it is perhaps not the ideal terminology for the relationship at hand. The implication is always that the equations specifying two rather different models (say, the model of a simple pendulum with small displacement and the model of an RLC circuit) are 'the same'. This does not mean the equations are *identical* however. In an RLC circuit, for example, the equation for the resonance frequency is $\omega = (LC)^{-\frac{1}{2}}$, where L is the inductance, and C is the capacitance. In a simple pendulum, $\omega = (Lg)^{-\frac{1}{2}}$ where L is the length and g is the acceleration of gravity. Those are not literally the same equation since one relates frequency to inductance and capacitance, etc. But we can easily see that they have the same structure – they are 'isomorphic' in syntax.

since, strictly speaking, the equations on either side of the isomorphism are not laws. To establish a full nomic connection we would have to relate the laws governing the fundamental dynamics of quantum phenomena at the horizon of a black hole (i.e. the relevant equation from a prospective theory of quantum gravity) to those governing the fluid flow (i.e. at the very least the full Navier-Stokes equations). A doubly infeasible task.

Nevertheless, there *is* a syntactic isomorphism to be exploited in the dumb holes case, and we think it is best understood as holding between two very particular *modelling frameworks*, each with narrower scope than genuine laws. The question is not of an isomorphism between the laws of fluids and the laws of quantum gravity on the other. Rather, there is an isomorphism between a particular adequate way of modelling a special class of fluid setups and a particular adequate way of modelling the behaviour of quantum fields near a black hole horizon.

With this in mind let us introduce some vocabulary concerning models. We build models using a *modelling framework*, M. The ideal pendulum is a modelling framework for modelling particular pendulums. The two-body, point-particle modelling framework is good for modelling planetary orbits under certain conditions. Modelling frameworks almost always involve idealizations, and hence they are usually adequate only under a certain *domain of conditions*, D. The domain of conditions under which the ideal pendulum framework is adequate will depend on what we mean by adequate. In general, adequate means 'for a particular purpose: accurate to a certain desired degree.' People who use and build models have lots of background knowledge, some of it explicit and some of it tacit, about what domain of conditions needs to apply before a particular modelling framework is adequate for a particular purpose and to a particular desired degree of accuracy.

Given the above, we will say that system S provides an analogue simulation of system T when the following set of conditions obtain:[7]

Step 1. For certain purposes and to a certain degree of desired accuracy, modelling framework M_S is adequate for modelling system S within a certain domain of conditions D_S.

Step 2. For certain purposes and to a certain degree of desired accuracy, modelling framework M_T is adequate for modelling system T within a certain domain of conditions D_T.

Step 3. There exists exploitable mathematical similarities between the structure of M_S and M_T sufficient to define a syntactic isomorphism robust within the domains D_S and D_T.

Step 4. We are interested in knowing something about the behaviour of a system T within the domain of conditions D_T, and to a degree of accuracy and for a purpose consistent with those specified in Step 2. For whatever reasons, however, we are unable to directly observe the behaviour of a system T in those conditions to the degree of accuracy we require.

Step 5. We are, on the other hand, able to study a system S after having put it under such conditions as will enable us to conclude a statement of the form:

Claim$_S$ Under conditions D_S and to degree of accuracy that will be needed below, we can for the purpose of employing the reasoning below assert that a system S will exhibit phenomena P_S.

The formal similarities mentioned in *Step 3* then allow us to reason from *Claim$_S$* to *Claim$_T$* which is of the form:

[7]The following is consistent with the account of simulation offered in (Winsberg, 2009; Winsberg, 2010)

Claim$_T$ Under conditions D_T a system T will exhibit phenomena P_T.

As a tool for analysing the structure of contemporary science this notion of simulation has the following advantages over the Hempelian category of nomic isomorphism. First, as we have already noted, the requirement that the syntactic isomorphism be between two sets of laws is too strong to cover most of the interesting cases. It seems clear that the kind of reasoning involved in dumb hole cases is more in the spirit of what Hempel had in mind than it is in the spirit of the sort of thing described by Keynes and Hesse, or of that exemplified by the Reid example. However, the dumb hole cases are difficult to see as fitting in to the strict set of requirements set out by Hempel. The relevant fact is not whether there is a formal relationship between two sets of laws, but rather whether such a relationship obtains between two suitably useful modelling frameworks.

Second, we think analogue simulation is better seen as distinct from analogical reasoning than as a sub-species of analogical reasoning. This is because the strength or quality of the inferences one can draw by analogue simulation is much greater than is that of those which can be drawn via analogical reasoning. We think analogue simulations can provide much stronger support for the conclusions we draw from them – this is of course the basis behind the key confirmation claims discussed at length in the following sections. Furthermore, we hope that our characterization of analogue simulation emphasizes the extent to which what we are dealing with is not simply a form of abstract argument, but rather a technique for learning about the world by manipulating it. It should be obvious, for example, that employing dumb holes to learn about Hawking radiation has a lot more in common with experiment than does Reid's armchair speculations about life on other planets. *Step 5*, above, after all, will require us to manipulate a system S so as to put it in the set of conditions that allow us to make the inference to [$Claim_T$].

Third, we find it attractive that it is an easy consequence of our characterization of analogue simulation that analogue simulation and computer simulation come out as two species of the same genus: simulation.[8] On our view, the main difference between computer simulation and analogue simulation is simply that in computer simulation, the system S is a programmable digital computer, and the reasons that it meets the conditions articulated in *Step 2* is that it has been programmed precisely so as to meet those conditions. The programmable digital computer is such a powerful scientific tool precisely because it can be so easily prepared in such a way. We find this consequence attractive, in part, because we find that it accords nicely with scientific practice and the intuitions of working scientists, who view the kind of work exemplified by dumb hole studies as easily comparable to computer simulations. We will comment further on the comparison with computer simulation in Sect. 5.3.3, once the major argument of the paper concerning confirmation has been introduced.

5.3.2 Confirmation via Analogue Simulation

The groundwork has now been laid for us to make our most controversial claim: that, in certain circumstances, analogue simulation can provide inductive support for a hypothesis regarding the target system, on the basis of empirical evidence regarding the source system – in other words, it can give us confirmation. It should be noted, that for the purpose of the present analysis we will not be concerned with the possibility of characterising cases of analogue simulation in terms of a particular

[8]See (Winsberg, 2010) for more on this point.

philosophical model of confirmation.[9] Rather, we will propose that certain cases of analogue simulation should plausibly be counted amongst the explananda for which the models of confirmation are intended to provide the explanans. From our perspective, if it proves that a philosophical model of confirmation cannot accommodate confirmation via analogue simulation at all, then this would be as much a problem for the model, as it would for analogue simulation.

That said, our aim here is emphatically not to propose a new category of confirmation based on analogue simulation merely on the basis of the intuitions and practices of 'working scientists'. Thus, we seek to carve a middle course between the normative and descriptive: neither assuming an abstract model of what should be counted by scientists as confirmatory, nor transcribing from their practice, a model of what actually is. We claim that philosophy of science is hostage to scientific practice and thus as the latter evolves, so must the former. That is, insofar as scientists develop novel methods for confirming hypotheses, our own models of confirmation must adapt to them. However, on our view, such novel practices must still be subjected to detailed philosophical analysis, and the cogency of claims of confirmation via novel methods tested all the more robustly on the grounds of their novelty. This is best done without prejudicing the analysis by adopting a particular account of confirmation. Only then, if the claim that a novel scientific practice is confirmatory survives such an analysis, can it be claimed that a philosophical model of confirmation needs to accommodate the new confirmatory mechanism.

[9]Major approaches to confirmation theory (according to a relatively standard classification) are: confirmation by instances, hypothetico-deductivism, and probabilistic (Bayesian) approaches. See (Crupi, 2013) for more details in general, and (Dizadji-Bahmani, Frigg, and Hartmann, 2011) for work on applying the Bayesian framework in the context of analogical relationships. A forthcoming paper ("Confirmation via Analogue Simulation: A Bayesian Account") will explore the foundations of analogue simulation from a Bayesian persecutive.

Things of course also cut the other way. Nothing in the proposed methodology rules out the possibility that the scientist's claims (regarding the novel methods and confirmation) might either partially or entirely fail to live up to careful philosophical scrutiny. Our analysis, in such a case, would then licence normative arguments *against* the scientific intuitions. What is more, it could also be used to criticise philosophical models of confirmation for being overly permissive, rather than excluding new phenomena. In what follows we will, in fact, detail the extent to which 'confirmation via analogue simulation' can fail to hold. Thus, our analysis will also serve as a basis to identify cases which *should not* count amongst the explananda of models of confirmation, and we believe it is entirely possible that the existing philosophical models might be troubled by accommodating analogue simulation too easily.

The natural starting point for the discussion of confirmation in our specific case is an argument given by Bartha (Bartha, 2013) against analogical reasoning being confirmatory. This argument is based upon a specific philosophical model of confirmation – Bayesian confirmation theory. However, in line with the considerations above, such a model will not be assumed in our positive story. From a Bayesian perspective it seems reasonable to assume that evidence for a hypothesis can count as confirmatory only if the probability of the hypothesis given the evidence together with certain background assumptions is larger than the probability of the hypothesis given only the background assumptions. Bartha contends that we should take the information encapsulated in an analogical argument to already be part of the background knowledge, and thus the probability of a hypothesis regarding the target system will be identical before and after finding any empirical evidence regarding the source system. It seems reasonable to accept this argument for the case in which the target and source are merely connected by analogical reasoning. However, when an analogical connection is established via analogue simulation there are good reasons to doubt the Bartha argument: prima facie, we do have the collection of new evidence – i.e. evidence which is not part of that background knowledge. Within our characterisation, this new evidence feature is found precisely when we probe the phenomenology of the source system to gain new evidence regarding an exemplar of the target system (i.e. *Step 5*). Of course, that this new evidence really is evidence relevant to the target system is only the case given the all important assumptions regarding the accuracy of the modelling frameworks and the robustness of the syntactic isomorphism (i.e. *Steps 1-3*). And clearly such assumptions are open to question without further support. Moreover, how can we be justified in thinking that the syntactic isomorphism will still hold within domains where we do not, by assumption, have any evidence that the modelling framework relevant to the target system is accurate?

Let us be more explicit regarding the required structure. The claim is that new evidence for the phenomena P_S as predicted by the model M_S of the system S, can confirm the existence of the analogous phenomena P_T, as predicted by the model M_T of target system T. For such a claim to be justified not only must equivalence be established between the mathematical descriptions of P_S and P_T, but we must also have evidence that both modelling frameworks will be accurate within the domain in which the phenomena are found. The problem of course is that it is not obvious how the accuracy of the framework M_T within the domain D_T could be established empirically with regard to the phenomena P_T, since we have assumed this phenomena to be inaccessible! Rather we need reasons external to the particular modelling frameworks at hand to justify the robustness of the formal correspondence.

We can see how this can be done as follows. Given certain explicit assumptions $\mathcal{A} = A_1, ..., A_n$ about the model M_T of target system T, we are able to derive P_T. These assumptions are based on some additional implicit assumptions $\mathcal{I} = I_1, ..., I_m$ of the sort 'property X of system T does not influence the derivation'. These implicit assumptions are not "premises" of the derivation but are, if true, the justifications for the use of the assumptions $\mathcal{A}$. Now system S is modelled in such a way as to realize the assumptions $\mathcal{A}' = A'_1, ..., A'_n$, in model M_S. These assumptions are syntactically isomorphic to the assumptions $\mathcal{A}$ in model M_T. Therefore the derivation of P_S goes through within model M_S as it did in M_T and we have a mathematical descriptions of the relevant phenomena which are suitably isomorphic. The set of underlying implicit assumptions $\mathcal{I}' = I'_1, ..., I'_k$ for model M_S can possibly be different from those of model M_T and no syntactic isomorphism between the set of assumptions are needed here since they are not used in the derivation.

With respect to system S one has control over the realization of the assumptions $\mathcal{A}'$ necessary for the derivation, i.e. the required properties can to some extent be realized by construct. However, if system T is inaccessible it remains an open question whether the assumptions are actually realized there. The reason for this is that the implicit assumptions $\mathcal{I}'$ justifying the assumptions $\mathcal{A}'$ in model M_S are different from those justifying the assumptions $\mathcal{A}$ in model M_T. And knowledge about the $\mathcal{I}'$s do not necessarily entail any information about the implicit assumptions in the other system, since a priori these are independent. That is, in the worst case, one models in system S something which is not realized in target system T.

So unless we have some reason to relate the implicit assumptions in both systems with each other there is no reason why observation of the phenomenon in one system should entail empirical evidence for the other. However, it is our claim, that in the case of analogue models of black hole Hawking radiation a relation between these implicit assumptions can be formulated which can be empirically tested and so build an empirical bridge between the target and source system.

Let us illustrate this with a simple example before discussing the dumbholes more closely in the next section. There is a syntactic isomorphism between Newton's Law of Gravity $F_N = G\frac{m_1 m_2}{r^2}$ and Coulomb's law $F_C = K\frac{q_1 q_2}{r^2}$.[10] Let us, for the purpose of the argument, assume that Newton's law describes our inaccessible target system while we are able to test Coulomb's law. First note, how analogue simulation can obtain between systems with radically different ontologies. In one case we consider the strength of the interaction between charged objects while in the other case it is the strength of the interaction between massive bodies. But why should the force law between massive and charged objects be syntactically isomorphic?

One crucial formal feature behind the syntactic isomorphism is the dependence of the force law on the distance r of the two bodies. In both cases we have a $1/r^2$-dependence. One assumption that goes into the derivation of both laws and which gives rise to this dependence is the dimensionality of space. One can show that if space were to be 2-dimensional the force law in both cases would go as $1/r$ while in four space dimensions the laws would go as $1/r^3$. It is only in three space dimensions that the $1/r^2$-relation obtains.[11] This illustrates nicely how

[10]This example serves the purpose of exemplifying some of the aspects of analogue simulation but can actually not be applied to our framework since Coulomb's law describes static interactions only and so could not be used to test e.g. planetary motion.

[11]Assuming the Poisson equation holds in all dimensions. There are also some other subtleties. See (Callender, 2005) for details.

specific features, the dimensionality of space, common to both systems can lead to a syntactic isomorphism despite them having radically different ontologies.

The existence of such common features also serves as a guide to the conditions for the accuracy of the modelling frameworks and robustness of the isomorphism within the relevant domains. It is, in fact, clear from modern particle physics that not all fundamental interactions follow a $1/r^2$-dependence in three spatial dimensions.[12] In particular the weak nuclear force does not. Formally such features can be understood in the framework of quantum field theory (QFT) as relating to the mass of the force mediating bosons: QFT interactions are always mediated via so-called gauge bosons and a $1/r^2$-dependence obtains only if we have massless-mediating particles. The gauge bosons of gravity and electromagnetism (the graviton and photon respectively) are massless, thus we get a $1/r^2$-dependence. On the other hand the mediating bosons for the weak forces (W and Z bosons) are massive, so the r-dependence is more complex.[13] The description of interactions via QFT is empirically well confirmed, and thus this argument towards the level of generality of the $1/r^2$-dependence is not merely theoretical. It gives *empirically grounded* and *model-external* arguments for the robustness of the syntactic isomorphism between the Newtonian and Coulomb modelling frameworks[14].

[12] There are physical proposals, like the Arkani-Hamed-Dvali-Dimopolous model (Arkani-Hamed, Dimopoulos, and Dvali, 1998) or the Randall-Sundrum model (Randall and Sundrum, 1999), which claim that there are additional dimensions. In the Randall-Sundrum model the claim that the dimensionality of space is a common feature of both gravitational and electromagnetic systems is denied: while electromagnetic interactions are restricted to three space dimensions, gravitational interactions are not. This shows that there is empirical room for these kind of modifications.

[13] For a simple treatment of this see Sec.1.4 and 1.5 in (Zee, 2010).

[14] One might object that quantum field theory applies, unlike Newton's law of gravity, to fields and not particles. But this misses the point, since the relevant question is whether the same phenomena, namely the $1/r^2$ relation in gravitational and electrostatic interactions are obtained. And since this is the case, QFT provides an independent justification for the robustness of the syntactic isomorphism.

Such additional knowledge coming from an underlying theory does not rule out the possibility of a breakdown in the syntactic isomorphism entirely, but it does give reason to insist the assumptions crucial to *Steps 1-3* above are well-founded.

We have thus seen that additional knowledge of the underlying physics can give us reason to believe in the correctness of the underlying implicit assumptions. In the example considered we had shared explicit assumptions of the sort that both systems follow the Poisson equation in three space dimensions (for example) and implicit assumptions that the very specific properties of each of these systems do not lead to a deviation from the law. More precisely, we implicitly assume that the way electric charges interact is similar enough to the interaction of massive objects so that the $1/r^2$-dependence in each case is robust. As we saw this assumption can be supported by *model-external* and *empirically grounded* arguments – let us abbreviate such arguments as 'MEEGA'. As we discussed above, the systems T and S differ in terms of the implicit assumptions. The non-realization of one of the implicit assumptions in the target system can lead to a failing of the analogue setup being able to confirm. MEEGA give us a handle on exactly these implicit assumptions and allow us to bridge in an empirically justifiable way the reasonableness of the implicit assumptions in the target system. Of course, this does not rule out the possibility that the violation of some unthought-of implicit assumption in the target system could lead to the target system not developing the phenomenon observed in the analogue system. Thus, MEEGA only ever give us an inductive base for believing in the robustness of the syntactic isomorphism and accuracy of the modelling frameworks, they do not establish anything tout court.

One might object that the establishment of MEEGA makes the need for the analogue simulation obsolete. This is not the case, since what MEEGA does is not to replace the modelling frameworks with an overarching theory but to justify empirically the validity of the implicit assumptions in each of the systems. As discussed above what a quantum field theoretic treatment is offering us is not a unifying theory of electromagnetic and gravitational interactions but constraints on any quantum field theoretical treatment of these interactions, thereby increasing our degree of believe with respect to the implicit assumptions and establishing a robustness of the modelling frameworks used. However, if there were one scientific theory that would cover both domains of applicability of the modelling frameworks, e.g. a unified theory of all fundamental forces, then, of course, the analogue simulation as a mean to confirmation becomes obsolete and we get back to standard theory confirmation by evidence.

So the lesson is, one is only justified in claiming confirmation via analogue simulation once one has established, via MEEGA, additional reasons for the accuracy of the modelling frameworks, and robustness of the syntactic isomorphism within the relevant domains. The key question examined in the remains of this paper is whether such conditions can be established for the case of dumb holes and black holes. However, before we enter into this discussion it will be instructive to review the crucial terminology and distinctions that have been introduced so far.

5.3.3 Recapitulation

There are three separate distinctions regarding types of scientific inference relevant to our argument. The first distinction is between ordinary analogical reasoning and analogue simulation and has already been discussed extensively in Sect. 5.3.1. The second distinction is between analogue simulation and computer simulation. Although analogical simulation is very much like computer simulation we do not here take them to be identical. There is a clear and obvious difference between the two in that in the case of computer simulation the structural similarity that exists between the model of the target and the model of the source exists precisely because the source system has been digitally programmed in such a way as to make this isomorphism obtain. However, there is a strong correspondence between such digital programming, and the preparation of the analogue model: the preparation of a fluid to formally resemble a black hole is, in a sense, a form of analogue programming (where here we are using 'analogue' in the 'not digital' sense of the word). We concede, therefore, that our claim that analogue simulation can confirm beliefs about the target simulation is not unrelated to the parallel claim about computer simulation, and we note that the latter claim is controversial in the literature. See (Beisbart and Norton, 2012) for a defence of the negative claim, and (Parker, 2009; Winsberg, 2009), for a defence of the positive claim.

As it so happens, we take the positive side of this debate: since programmed digital computers are physical systems, a run of such a system gathers novel empirical evidence, and so surely must in principle be able to provide confirmation. Thus, on our view, and contra (Beisbart and Norton, 2012), computer simulation is not simply an 'argument', and can *in principle* boost our degree of belief in a hypothesis about the target system, provided that the relevant background knowledge (concerning the isomorphism) is in place to support the relevant inference. Of course the run of a computer simulation will never serve to confirm all of the background knowledge supporting the claim of an existing structural similarity between source and target. A computer simulation of a fluid that uses the Navier-Stokes equations to guide its construction will never, by itself, confirm the Navier-Stokes equations themselves. But that does not mean that such a computer simulation cannot confirm, e.g., a scaling law regarding certain kinds of fluid configurations.[15]

This argument of course depends crucially upon the inference from the premise that programmed digital computers are physical systems, to the conclusion that a run of such a system can gather novel empirical evidence. Since, in such cases, the novel empirical evidence comes from a physical system whose job it is to perform calculations that could, in principle, have been carried out by rote on a piece of paper, a critic might – in support of (Beisbart and Norton, 2012) – argue that we do not have *genuinely* novel empirical evidence in the sense that matters to confirmation theory. We believe that even if one accepts critical arguments in this vein, there are still cases of computer simulation that should plausibly be counted as confirmatory. And the reason why is, in fact, closely related to the idea of analogue simulation supported by model-external and empirically grounded arguments (i.e., MEEGA).

[15]To give a concrete example, it is widely held by astrophysicists that certain computer simulations carried out in the early 1970s (see for example (Toomre and Toomre, 1972)), confirmed the previously heretical claim that tidal forces that arise when two galaxies collide were responsible for the phenomenon of galactic tails and bridges. We see no reason to deny this, since novel evidence came from the runs of of these simulations that rationally raised people's degrees of belief in the hypothesis. But of course, no one, on the other hand, would claim that these simulations could confirm the existence of tidal forces.

In practice much interesting science involves not just computer simulation in isolation, but the gathering of evidence regarding the reliability of the syntactic isomorphisms that such simulations exploit. Well-designed computer simulation studies generally involve a back-and-forth between simulation runs and real-world data gathering in support of their own background assumptions. Thus, certain cases of computer simulations involve the activity of providing MEEGA for the background assumptions used in other computer simulation, including, in some cases, themselves. In such cases, the role of simulation explicitly could not, even in principle, be played by calculations on a piece of paper, and thus the criticism of the evidence as not *genuinely* novel, falls short.

Thus, we do, as a matter of fact, think it is very plausible to believe that computer simulations can confirm certain hypotheses. However, this is not a central claim of this paper. *One can reject the arguments in favour of computer simulations being confirmatory, but still accept that dumb hole experiments can confirm the existence of Hawking radiation.* This is because of the third, and most important, distinction: the distinction between generic analogue simulations, on the one hand, and analogue simulation supported by MEEGA, on the other.

One might suspect that the arguments against computer simulation being confirmatory might be applicable to a general case of analogue simulation. However, as we have seen, such arguments are much less plausibly applicable in the case that computer simulations are engaged in, or supported by MEEGA. Similarly, for the case analogue simulation supported by MEEGA, the critical argument against the collection of novel evidence based upon the comparison with pen and paper calculation surely must fail entirely. In the case of an analogue simulation supported by MEEGA, by definition, there are model external, empirically grounded reasons to believe that novel phenomenology is being simulated. Thus, the collection of novel evidence is secured in a sense much stronger than that of either generic computer simulation or analogue simulation.

A further interesting point relates to the notion, mentioned above, of simulations which provide their own MEEGA. Such an idea will, in fact, prove to be embodied in precisely the case of analogue simulation under consideration. The central concern in this paper is the chain of arguments by which one might reasonably claim that dumb hole studies could confirm the existence of (gravitational) Hawking radiation. On our view, confirmation of Hawking radiation via analogue simulation can only be established given the acceptance of a chain of reasoning involving universality arguments in combination with diverse realisations of the counterpart effect. These diverse realizations will thus simultaneously provide the empirical support for the MEEGA supporting the simulation, *and* realize the simulations themselves. We will return to this point in Sect. 5.4.2.

5.4 The Sound of Silence: Analogical Insights into Gravity

5.4.1 Experimental Realisation of Analogue Models

The wish to use analogue models to test inaccessible target systems did not grew out of sheer creativity of the scientists involved but out of necessity. Science has reached a point where many theoretical ideas can not be tested due to several practical limitations. These limitations can have several reasons. For instance, the technology to test the theory has not yet been developed, the system that needs to be tested is unreachable, there is simply not enough funding to build the experiment, or a combination of the above. However, practical limitations to test a theory do not make a theory less scientific, and therefore the question remains of how then to establish confidence in these theories. In this context, analogue simulations have proved as a promising alternative and several applications beyond the dumb hole case, whose experimental realisation we will discuss now, have been proposed. Some further applications will be discussed in Sect. 5.5.

If the analogue models of black hole physics were purely hypothetical then their scientific status would likely be merely that of a fascinating novelty, rather than the inspiration for an entire sub-field of modern physics. After all, the main problem with theoretical work concerning black holes is that we have, as yet, no empirical means of testing the predictions – since we can neither create black holes in the lab, nor probe them via astrophysical observation. In the spirit of our view of the dumb hole model as a simulation of a black hole, the situation, if one were *not* able to experimentally realise the model, would be like having the correct code for a computer simulation, but not being able to run it due to unrealistic hardware requirements. A not particularly useful situation. Here we will briefly survey the practical problem of detecting Hawking radiation via analogue simulation, and in doing so consider further models beyond the fluid mechanical dumb hole discussed above.

First let us consider a model along the lines of Sect. 5.2.3 with the flowing through a nozzle in order to create the deserved acoustic horizon. Following, (Novello, Visser, and Volovik, 2002, p.26) we have that for supersonic flow of a fluid through a nozzle of radius R, the approximate value of T_H^{acoustic} is given by the expression:

$$T_H^{\text{acoustic}} = 1.2 \times 10^{-6} K \left[\frac{c_{\text{sound}}}{1km/s} \right] \left[\frac{1mm}{R} \right] \tag{5.14}$$

For water this equates to a temperature of the order 10^{-6} K. Detecting a thermal phonon spectrum at this temperature while the ambient temperature is approximately 300 K is entirely impractical, and so clearly water is not going to provide a useful working fluid for real laboratory experiments of this form. However, there are other methodologies for setting up the analogue model in which the use of conventional fluids for detecting at least *classical aspects* of Hawking radiation becomes more practicable (Weinfurtner et al., 2011; Weinfurtner et al., 2013).

These experiments are based on a proposal by Schützhold and Unruh (Schützhold and Unruh, 2002) where surface gravity waves in water tanks are used instead of Unruh's original proposal of sound waves. The problem for the sound wave proposal is that the acceleration of fluids to velocities close to the speed of sound leads to turbulences due to shock waves. Once there are turbulences the linearisation assumptions involved in the derivation of the effect are not realised anymore and experimental control of these turbulences is too difficult. The new proposal by Schützhold and Unruh considers surface waves on shallow water flows. The advantage is that the velocity of the background flow v and the velocity of the surface waves c are both dependent on the depth h of the water tanks. More concretely, the velocity of the background flow goes as $v \propto \frac{1}{h}$, while the surface waves velocity goes as $c \propto \sqrt{h}$. So by adding an obstacle to the water flow, h decreases, and thereby the background flow velocity increases while the surface wave velocity decreases. If the velocity of the background flow exceeds the velocity of the surface waves an effective horizon is obtained.[16] The surface waves are effectively 'blocked' by the horizon, so nothing can enter the critical region. In this sense the experiment is a fluid mechanical analogue of a white hole, i.e. the time inverse of a black hole, for which the same laws hold.

The results of the experiment detailed in (Weinfurtner et al., 2013) imply that the pair-wave creation at the effective white hole horizon has a thermal spectra consistent with that predicted by the Hawking type calculation: the generated incoming positive mode is converted at the white hole horizon to positive and negative outgoing modes. And the corresponding amplitudes of the outgoing modes, i.e. the Bogoliubov coefficients, were measured and the validity of the thermal spectra according to Hawking checked. The dispersion relations used are different from the ones assumed by Hawking in his derivation which leads the authors of the paper to the following conclusion (p.15):

> The ratio is thermal despite the different dispersion relation from that used by Hawking in his black hole derivation. This increases our trust in the ultraviolet independence of the effect, and our belief that the effect depends only on the low frequency, long wavelength aspects of the physics.

We will comment more on such issue regarding the relevance or not of short wavelength physics shortly. There are two drawbacks of this experiment. First, the experiment considers only the stimulated Hawking Process and not the spontaneous. Second, only classical aspects of the process are measured in the experiment. The behaviour for this linear quantum system is dominated by the classical behaviour and additional

[16] The velocities needed to develop an effective horizon are low and the avoidance of turbulence is experimentally realisable and is being tested in the experiments.

quantum correlations of the emitted field excitations are not measured by the experiment. However, other analogue experiments based on Bose-Einstein condensates may be able to take the analogue experiments to the quantum regime.

The fluid mechanical analogy has proved to be only one among a number of possible realisations for the Hawking phenomena in terms of small oscillations in continuum systems. In general, it seems that there are only two necessary requirements to reproduce Hawking radiation (Barceló, Liberati, and Visser, 2005), in that we need: i) a quantum analogue model with a classical effective background and relativistic quantum fields living on it; which ii) contains some analogue geometry with some sort of horizon. Thus far, in addition to sound in a liquid and surface waves in water tanks, contemporary analogue gravity research (both theoretical and practical) makes use of: phonons in superfluid liquid helium, atomic Bose-Einstein condensates (BECs) (Garay et al., 2000) and degenerate Fermi gases; 'slow light' in moving media; and traveling refractive index interfaces in nonlinear optical media (see (Carusotto et al., 2008) for further references).

5.4.2 Universality and the Hawking Effect

We must now turn the focus of our analysis to a problem of particular importance for both the reliability of the calculation of the Hawking effect, and our claim that the relevant analogue gravity models should be understood in terms of the concept of analogue simulation. In the standard calculation of the Hawking temperature, which is used in both the gravity and analogue cases, the black hole radiation (or its analogue) detected at late times (i.e. the outgoing particles) must be taken to correspond to extremely high frequency radiation at the horizon. This is because of an exponential gravitational red-shift (or its analogue) that is assumed to take place near the horizon. The problem with this is that the frequencies in question can in fact be high enough to make the relevant length scales smaller than those upon which the semi-classical approximation made in theories being applied are expected to work. For gravitation this corresponds to Planck-scale lengths at which it is no longer reasonable to use quantum fields defined upon fixed classical spacetime backgrounds. Such a 'trans-Planckian' regime is the dominion of theories of quantum gravity, and is thus well beyond the domain of applicability of the modelling framework we are using. This problem with trans-Planckian modes has a direct analogue in the fluids case, where the breakdown due to neglected quantum gravity effects is paralleled by that due to the atomic nature of the fluid. Thus, in each case the modelling framework we are using is in fact, strictly speaking, being applied beyond its proper domain of applicability. In and of itself, this is clearly a severe problem for the reliability of the Hawking radiation calculation in both the gravitational and the analogical situations. There is a sense in which such models fail the seemingly fundamental test of self-consistency.

Furthermore, in light of the trans-Planckian problem (which we will use as a generic name for the breakdown of both gravitational and analogue models at small distances), it becomes questionable whether the notion of analogue simulation as we have defined it is really appropriate. Specifically, given these issues it seems reasonable to worry that both *Steps 1 and 2* should be seen to fail, and thus that we have inappropriately applied our own concept of analogue simulation! Rather, perhaps, the appropriate philosophical framework for dumb hole models *is* something like analogical reasoning, as traditionally conceived. And for this case at least, our efforts in introducing the new conceptual framework of analogue simulation have been in vain.

Fortunately, things are not quite as bad as they seem. As we shall see, although not entirely solved, the trans-Planckian problem can be reformulated such that we can give precise conditions under which Hawking effect calculations are reliable. The relevant definitions of the modelling frameworks and domains of applicability can then be appropriately amended, and both Hawking radiation and analogue simulation can be saved from the trans-Planckian spectre. Before we consider such more sophisticated arguments regarding the relevance of trans-Planckian effects, it will prove particularly interesting, from an analogical reasoning perspective, to first consider the chain of theoretical developments that lead to the formal arguments that will be presented. Let us start with a key observation by Unruh regarding the differing epistemological statuses of the two breakdowns:

> At wavelengths shorter than the inter-atomic spacing, sound waves do not exist and thus the naive derivation of the temperature of dumb holes will fail. But unlike for black holes, for dumb holes, the theory of physics at short wavelengths, the atomic theory of matter, is well established. For black holes, a quantum theory of gravity is still a dream. Thus, if one could show that for dumb holes the existence of the changes in the theory at short wavelengths did not destroy the existence of thermal radiation from a dumb hole, one would have far more faith that whatever changes in the theory quantum gravity created, whatever nonlinearities quantum gravity introduced into the theory, the prediction of the thermal radiation from black holes was robust. (Unruh, 2008, p.2908)

This is of course a beautiful example of analogical reasoning, perhaps more attractive, but of the same genus to that of Reid. If we could show that the trans-Planckian fluid dynamical effects are irrelevant in our model of the fluid, we could speculate the same may be true with the gravitational case also. Even if we cannot make use of a precise mathematical relationship between the two situations to make an inferences à la analogue simulation, we could use the knowledge that they are similar in certain respects to make a less reliable inference à la analogical reasoning.

A specific suggestion regarding modelling of the breakdown for the fluid mechanical case was in fact made by Jacobson (Jacobson, 1991; Jacobson, 1993) in the early nineties: one can focus upon the altered dispersion relation (i.e. relationship between frequency and wavenumber) that is relevant to an atomic fluid rather than continuous fluid, and consider whether, in such models, the exponential relationship actually holds between the outgoing wave at some time after the formation of the horizon, and the wavenumber of the wave packet (Unruh, 2008). Approximate answers to such questions can be determined in practice via numerical methods, and it was shown by Unruh in 1995 that the altered dispersion relation in atomic fluids does imply that the early time quantum fluctuations that cause the late-time radiation are not in fact exponentially large (Unruh, 1995). Thus, there is sufficient basis to *analogically reason* that the trans-Planckian alterations to the gravitational dispersion relationship might also prove irrelevant to the Hawking effect. However, the situation has in fact proved much better than this: we can reasonably establish the applicability of full analogue simulation despite trans-Planckian effects in both systems. Recent work in fact allows us to generalise from the specific fluid dynamical alterations to the dispersion relation, to a model with a *generically altered* relation, independent of the particular cause of the trans-Planckian breakdown. Of particular relevance are calculations to this end by Unruh and Schützhold (Unruh and Schützhold, 2005). Their results represent a generalisation of earlier work by Corley (Corley, 1998) and provide a basis for a universality claim with regard to the Hawking effect.[17] Unruh and Schützhold demonstrate that specific assumptions and approximations can be made such that the role of possible trans-Planckian effects is factored into the calculation of Hawking radiation, and that, for such cases, the additional effects are found not to disturb the thermal spectrum as originally derived by Hawking. The Hawking effect does not, to lowest order, depend on the details of the dispersion relation at high wave numbers given certain modelling assumptions.

What is more desirable, however, is a set of general conditions under which such effective decoupling between the sub- and trans- Planckian physics can be argued to take place. Unruh and Schützhold's proposal in this vein runs as follows: we assume the breakdown of geometric optics (leading to the creation of particles) occurs in the vicinity of the horizon only. There the gravitational redshift induces a transition of trans-Planckian into sub-Planckian modes. Given this assumption, if the modes leave the Planckian regime in their ground state with respect to freely falling observers[18] near the horizon, then Hawking radiation can be obtained (Unruh and Schützhold, 2005, p.8). The condition of the modes leaving the Planckian regime in their ground state with respect to freely falling observers is then taken to obtain in circumstances which: a) there is a privileged freely falling frame (in line with the breaking of

[17]For further work on these issues, using a range of different methodologies, see for example (Himemoto and Tanaka, 2000; Barceló, Garay, and Jannes, 2009; Coutant, Parentani, and Finazzi, 2012).

[18]For the dumb hole case the freely falling frame corresponds to the local rest frame of the fluid flow.

Lorentz invariance at the Planck scale); b) the Planckian modes start off in their ground state; and c) the evolution of the modes is adiabatic and therefore the Planckian dynamics is understood to be much faster then all external variations.

We thus see that there are good *theoretical* reasons for a qualified claim that Hawking radiation is a universal effect. Such claims provide a model-external basis for both the accuracy of the modelling frameworks and the robustness of the syntactic isomorphism within the domains of the gravitational and fluid mechanical Hawking phenomena. However, in order for such external lines of reasoning to count as genuine justification we earlier insisted that they be empirically grounded. How can we do this for our universality arguments? The answer is quite simple: we can empirically ground the universality argument vis-à-vis gravitational and fluid mechanical Hawking radiation, by empirically grounding *the same* argument as applied to other pairs of analogue models. Replication of Hawking phenomena could be achieved experimentally in systems as different as phonons in superfluid liquid helium, or traveling refractive index interfaces in nonlinear optical media. The fact that each system has a different underlying microphysics then gives empirical reasons to support the universality claim. This claim then justifies assuming the accuracy of the modelling frameworks and robustness of the syntactic isomorphism for the gravitational-acoustic case.

It is important to realize that the universality claim is not a theory replacing the modelling frameworks themselves. It is not a theory of slow light in dielectric media, sound in fluids and quantum fields around a black hole horizon at the same time. The sole purpose of the universality claim is to establish the independence of the Hawking radiation derivation in all these different systems from possible high energy physics effects, whatever these may be. And so, once it is empirically grounded, it establishes the validity of the most crucial implicit assumption in the derivation of the phenomenon.

5.4.3 Confirmation of Gravitational Hawking Radiation

All the features of our argument towards the possibility of confirmation of gravitational Hawking radiation via analogue simulation are now assembled. We are now in the position to make the central, and we believe rather bold, claim of the paper: Current and future experimental evidence of Hawking radiation from dumb holes can provide a high degree of warrant to claims regarding the existence of the phenomena in gravitational systems. Confirmation in the one case, can be understood as constituting confirmation in the other. We can proceed towards the establishment of this claim as follows:

Step 1. For certain purposes and to a certain degree of desired accuracy, modelling framework M_S is adequate for modelling systems of type S within a certain domain of conditions D_S.

In our case, a system of type S will be a fluid flow set up. The modelling framework M_S will consist of treating a fluid as a continuous, compressible, inviscid medium, without relativistic effect, and under the conditions of being barotropic and locally irrotational, along with the three conditions mentioned in Sect. 5.2.3 – and then quantizing the linearized fluctuations. The domain of conditions D_S under which we consider M_S to be adequate for modelling systems of type S for the purpose of calculating an acoustic Hawking temperature high enough to be detectable under reasonable laboratory conditions are the ones in which the fluid is set up with a flow so as to have an 'acoustic horizon' of the various kinds we talked about in Section 4.1. The work in dealing with the trans-Plankian problem discussed above is sufficient to demonstrate the viability of the modelling framework for some further refinement of D_S.

Step 2. *For certain purposes and to a certain degree of desired accuracy, modelling framework M_T is adequate for modelling systems of type T within a certain domain of conditions D_T.*

The system of type T in this case is of course the astrophysical black hole. The modelling framework M_T is a semi-classical model for gravity in which we have: i) a fixed classical spacetime that features the establishment of an event horizon via gravitational collapse leading to a black hole; and ii) a quantum scalar field evaluated in the regions of past and future null infinity which are assumed to be Minkowskian. The domain of conditions D_T is limited to the times after the collapse phase of the black hole, the details of which are assumed to be irrelevant. The work in dealing with the trans-Plankian problem discussed above is sufficient to demonstrate the viability of the modelling framework for some further refinement of D_S. For example the conditions a-c proposed by Unruh and Schützhold and discussed above.

Step 3. *There exists exploitable mathematical similarities between the structure of M_S and M_T sufficient to define a syntactic isomorphism robust within the domains D_S and D_T.*

This step essentially follows from the two steps above given the refinements of D_S and D_T regarding the trans-Plankian issue. Given both theoretical arguments towards universality of Hawking radiation (which do exist) and empirical support for these arguments (which currently does not), we have justification for assuming the accuracy of the modelling frameworks within the relevant domains, and, furthermore, robustness of the syntactic isomorphism.

Step 4. *We are interested in knowing something about the behaviour of a system of type T within the domain of conditions D_T, and to a degree of accuracy and for a purpose consistent with those specified in Step 2. For whatever reasons, however, we are unable to directly observe the behaviour of an exemplar of a system T in those conditions to the degree of accuracy we require.*

We want to know whether black holes exhibit Hawking radiation (presumably to a degree of accuracy that is sufficient for the purpose of warranting various speculations about black hole thermodynamics, etc.) but we can neither create black holes in the lab, nor probe them via astrophysical observation.

Step 5. *We are, on the other hand, able to study an exemplar of a system of type S after having put it under such conditions as will enable us to conclude a statement of the form:*

Claim$_S$ *Under conditions D_S and to degree of accuracy that will be needed below, we can for the purpose of employing the reasoning below assert that a system of type S will exhibit phenomena P_S.*

We build an acoustic system of type S that meets conditions D_S and we successfully experimentally measure (confirming it to be genuine signal rather than noise) the acoustic Hawking Radiation predicted by M_S.

The formal similarities mentioned in *Step 3* then allow us to reason from the existence of acoustic Hawking Radiation in dumb holes to Hawking radiation in black holes. From this we progress to our key claim: confirmation via observation under conditions D_S of the existence of acoustic Hawking Radiation would allow us to speak of having confirmed the existence of the analogue phenomenon, gravitational Hawking Radiation. Mutatis mutandis for disconfirmation. Thus we see that so far as Hawking radiation goes, and given some external support via MEEGA, dumb holes have the potential to tell us rather a lot about gravity.

5.5 Prospectus

The dumb hole/black hole case is a powerful illustration of the relevance of the notion of analogue simulation to modern science, and we believe there is a large range of further possible applications of the idea. A non-exhaustivity list of prospective cases is: 1) use of the AdS/CFT correspondence to make predictions regarding quark-gluon plasma, super conductors and super fluids (Hartnoll, Herzog, and Horowitz, 2008; Faulkner et al., 2010); 2) use of two-component Bose-Einstein condensates to study cosmic inflation (Fischer and Schützhold, 2004); 3) use of trapped-ions to simulate neutrino oscillation (Noh, Rodríguez-Lara, and Angelakis, 2012); 4) there are many applications of analogue simulation in condensed matter systems due to difficulties in solving quantum many-body problems, e.g. simulating high-temperature superconductors through quantum-dot arrays (Manousakis, 2002), quantum phase transitions from a superfluid to a Mott insulator phase through trapped atoms in an optical lattice (Greiner et al., 2002) and many more[19].

Amongst the field of such potential examples, what seems the most straight forward candidate for analogue simulation in contemporary science is the simulation of one quantum system by another. This general idea goes back to Feynman (Feynman, 1982; Feynman, 1986), and has rather diverse applications in terms of both quantum simulation via a programmable quantum computer[20] and via a quantum analogue system.[21] Focussing on the latter, the essential idea is to manipulate a well controlled quantum system – such as atoms in an optical trap – such that the Hamiltonian evolution of a different system is implemented. One then aims to probe the properties of the relevant Hamiltonian, and in doing so produce experimental measurements of phenomena whose correlates in the target system are experimentally inaccessible.

A particularly impressive recent example is the simulation of 'Zitterbewegung' (trembling motion) phenomena long known to be predicted by the Dirac equation (Schrödinger, 1930), using a single trapped ion set to behave as an analogue to a free relativistic quantum particle (Gerritsma et al., 2010). As is noted by the relevant authors, this case bears a strong resemblance to the dumb hole simulation of Hawking radiation. It is also a good illustration of the idea of analogue simulation that we have articulated: we have two theories, non-relativistic quantum mechanics, according to the Schrödinger equation, and relativistic quantum mechanics, according to the Dirac equation. And we have an experimentally inaccessible novel phenomena, Zitterbewegung, which is predicted by

[19]See (Georgescu, Ashhab, and Nori, 2014) for a recent review.

[20]See (Deutsch, 1985; Bernstein and Vazirani, 1993; Shor, 1997) for key historical developments, and see (Nielsen and Chuang, 2010; Timpson, 2013) for general reference.

[21]See (Lloyd, 1996; Cirac and Zoller, 2012; Greiner et al., 2002; Ma et al., 2011) for more details of both the concept and its applications.

one theory. Then, in certain very specific circumstances, we have an isomorphism between the two theories which allows us to experimentally simulate certain relevant formal structures, the Dirac equation in 1+1 dimensions, and thus reproduce a correlate of the novel phenomena. Given isolation of suitable MEEGA, confirmation via analogue simulation could then be understood to function as with the Hawking radiation case. More detailed investigation of this and other cases will be conducted in future work.

Chapter 6

Conclusion

The lack of empirical data in some parts of fundamental physics made it necessary to consider alternative methods of theory assessment and their viability. A very useful concept for that purpose is the space of all theories. When we test theories, we are constraining theory space. The gold standard for assessing theory space has been and continues to be empirical data. But once we understand empirical testing as putting constraints on theory space, we open up other paths for theory confirmation. In principle, any evidence of any kind that can to even the slightest extent constrain theory space will provide us with a potential confirmatory strategy. Similarly, recognising how constrained theory space is, may provide us with useful information regarding theory development.

In this dissertation we have considered some of these relations between theory assessment and development on the one hand and theory space on the other. Let us summarise our results.

The concept of limitations on scientific underdetermination provides the basis for non-empirical theory assessment. Dawid's proposed arguments for limitations on scientific underdetermination are indirect methods to assess theory space. Indirect, in the sense that it is not theory space itself that is being explored. In Chapter 2, I consider the possibility to assess theory space itself. This leads to several problems. The theoretical problem is the problem that scientists in assessing theory space rely on both theoretical constraints as well as the legitimacy of the problem posed. Both of these are problematic as they do not solely rely on empirically confirmed aspects and can therefore lead to inadequate constraints on theory space. The structure problem is the problem that scientists in formulating theories need to represent their physical assumptions within certain mathematical structures. These structures, however, may not be unique and so may lead to a wrong assessment of theory space. The final problem that is considered is the data problem. Here it is argued that even the assumed to be certain empirical constraints may pose too strong constraints on theory space. We argued that if non-empirical theory assessment wants to provide a reliable method of theory assessment, these problems need to be confronted heads on. This, however, makes a change in scientific practice necessary.

Once we have considered the problems that arise in assessing theory space directly, we turned in Chapter 3 to the indirect arguments proposed by Dawid. More specifically the claim made in (Dawid, Hartmann, and Sprenger, 2015) that the No Alternatives Argument can confirm a theory. In the light of the results of Chapter 2 we first show that the non-empirical evidence used is insufficient to be applied to most circumstances of interest, and therefore needs to be complemented. Once complemented, however, its application to the case of quantum gravity becomes difficult. I argue that at this point in the history, the required constraints on theory space are not yet established. One reason for that is that different research traditions meet in their aim to unify quantum mechanics and gravity, possibly leading to conflicting constraints on theory space. As there is no agreement among scientists regarding the constraints on theories of quantum gravity at this point the application of the No Alternatives Argument is premature.

In the next chapter we switch from confirmation to theory development. No-go theorems are important methodological tools in theory development. They intend to show the impossibility of a certain goal and as such are explicit statements about theory space. They have played the role of stopping whole research programmes. But are these implications justified? In Chapter 4 I start by discussing an exemplar set of no-go theorems, which allows us to give an abstract definition of no-go theorems. Recognising this more abstract structure has several implications. For one, the no-go result is more complex than is usually assumed in the interpretations of no-go theorems and this has obvious implications for its interpretation. Given the more extended set of elements on which the modus tollens is applied, the interpretation depends strongly on the preference assignment on these elements. These are, however, not unique. So in general there is not a unique interpretation of a no-go result. In addition, I argue, that a problematic element of any no-go result is the mathematical structure. Its denial, however, together with the denial of the physical assumptions, imply methodological pathways in exploring theory space rather than the impossibility of any specific element. Thus, no-go theorems are best understood as go theorems.

In the final Chapter 5, we explored another possible path in assessing theories in the absence of direct empirical evidence. We assess the viability to use the experimental results of table-top analogue models of black holes to confirm aspects of the inaccessible target system, the black hole Hawking radiation. We call this analogue simulation. Analogue simulation is based on the idea that one simulates the equations of the inaccessible target system, by constructing an accessible model in such a way that it satisfies the same equations. We argue that (i), if we have no further information about the two systems we are not allowed to draw an inference from one system to the other. The reason for that is that what justifies the assumptions in the modelling framework in one system is independent from what justifies the assumptions in the other framework. What can make one model wrong, does not necessarily make the other model wrong. We argue, however, that in cases where these assumptions are related via model external and empirically grounded arguments, evidence in one system can confirm the other. This again has an obvious link to theory space, because whenever we rule out a possible way a theory or model could be wrong, as we do in analogue simulation, we constrain theory space by ruling out theories, which would have the feature of denying that very assumption.

While the title of this dissertation is "Challenging Scientific Methodology" an equally adequate title could have been "Exploring Theory Space". It is only through the concept of theory space and its exploration that we have been able to challenge the existing empirical paradigm in scientific methodology. The previous chapters have presented some of the possibilities that an explicit appeal to theory space allows.

Theory space plays a crucial role in many other fields in philosophy of science. Maybe, most importantly in the realism/anti-realism debate, where the underdetermination of theory by data is a strong argument against scientific realism. The argument by underdetermination is implicitly referring to a space of all theories, within which one assumes there are many theories able to account for the same data. That assumption is usually based on a logical/formal argument or on a historical inductive argument. None of these, however, assume that one can actually go about assessing theory space. It is always this not clearly defined abstract structure, within which philosophical speculation feels most comfortable in. But once we make theory space concrete, an object that needs to be studied explicitly, it may very well turn out, that the historical examples considered evidence for a general claim may just be contingent and that the formal/logical arguments provided are actually not talking about the relevant theory space.

A more explicit assessment of theory space will be crucial for an adequate development of the following issues, which need to be addressed:

- Scientific Methodology: There are many further issues that need to be addressed in the context of scientific methodology. To name a few. Analogue simulation needs to be situated on the 'methodological map' by delineating its methodological similarities and differences to analogical argument, experimentation, computer simulation, and quantum simulation. Can computer simulations confirm? Are experiments always more reliable than e.g. simulations of some kind? Similarly, we need to develop non-empirical theory assessment further. For one, it requires more detailed case studies in order to investigate its limitations and capabilities. More systematic study is needed as well to see how exactly the No Alternatives argument is related to the other arguments Dawid proposes. Furthermore, one needs to consider possible alternative approaches in assessing theory space.

- Epistemology: The question about the limits of scientific knowledge and similarly the question of how far empirical data can take us finds a different perspective within a theory space approach and may therefore provide us with interesting new insights. I believe exploring these questions within an explicit theory space will shed some light on the existing literature on this topic. Furthermore, a more detailed and unified Bayesian account of theory space exploration is needed and a related explication of confirmation and explanation within it.

- Metaphysics: As already mentioned, theory space plays an implicit role in much of the debates about the ontology of physics. The pessimistic meta induction argument, the no miracles argument and obviously the underdetermination argument rely crucially on a notion of theory space, which, however, remains implicit. Much of what we have discussed for instance in Chapter 2 can already be seen as possible problems for ontic structural realists.

- Other Sciences: Throughout this work we had physics in mind. However, there is no principled reason why the concept of theory space cannot be useful in either other mathematised sciences or even non-mathematised sciences. Mathematical modelling in the climate sciences as well as evolutionary models in the life sciences are examples at hand, where the concept can potentially be fruitful.

We have presented a new perspective on theory development and theory assessment based on taking theory space seriously. Considering different features of theory space allowed us to address many different questions. This was already possible without clearly defining theory space itself. Recognising some of its features turned out to be sufficient for the analysis provided here. But what is theory space and what is its ontological status? These are difficult questions with wide ranging consequences. These remain questions for another time.

Acknowledgements

First of all I would like to express my sincere gratitude to my supervisor Stephan Hartmann for his support and patient guidance throughout my doctoral studies. This document as well as my academic development in general benefited from my interactions with him in conversations, comments and advice.

There is not a single chapter in this dissertation, which was not influenced by Richard Dawid's work. So I am very grateful to him for agreeing to co-supervise me. Most chapters grew out or were significantly shaped by discussions I had with him and belong to the most intellectually enjoyable experiences I had during my studies.

It is a pleasure to thank all the people I have collaborated with in the last couple of years: Karim Thébault, Eric Winsberg, Stephan Hartmann, Luke Glynn and Mathias Frisch. My special thanks goes to Karim Thébault for his unweary enthusiasm in developing new ideas and projects.

I also owe a great deal to Mike Cuffaro, Erik Curiel, Sebastian Lutz and Karim Thébault for providing extensive comments and suggestions on many previous drafts. Their insightful remarks and criticism helped me develop my ideas to where they are now.

The MCMP is a remarkable place with an intellectual environment I have not experienced anywhere else. My thanks to Hannes Leitgeb and Stephan Hartmann who have made it such a wonderful place to work in. Many of the MCMP family have contributed in one way or another to the development of my dissertation: Alexander Reutlinger, Gregory Wheeler, Seamus Bradley, Maria Csauscher, Michael Stöltzner, Lorenzo Casini, Milena Ivanova, Sam Fletcher, Brian Padden and many others. I also like to thank the Alexander von Humboldt foundation for funding my doctoral studies at the MCMP.

During the last six months of my doctoral studies I had the great pleasure to work as part of the Geneva Symmetry Group. My thanks to Chris Wüthrich, Augustin Baas, Juliusz Doboszewski and Niels Linnemann for a very stimulating time and the John Templeton Foundation for funding my stay in Geneva.

I would like to thank my parents for their constant support and encouragement. Thanks also to my brother Mani for proofreading parts of the dissertation.

Finally, and most importantly, I want to thank my wife. It was only through her support and sacrifices that this work became possible.

Bibliography

Afriat, Alexander (2013). “Weyl’s Gauge Argument”. In: *Foundations of physics* 43.5, pp. 699–705.

Andersen, Hanne and Brian Hepburn (2015). “Scientific Method”. In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Winter 2015.

Arkani-Hamed, Nima, Savas Dimopoulos, and Gia Dvali (1998). “The Hierarchy Problem and New Dimensions at a Millimeter”. In: *Physics Letters B* **429**.3, pp. 263–72.

Barceló, Carlos, Luis Javier Garay, and Gil Jannes (2009). “Sensitivity of Hawking Radiation to Superluminal Dispersion Relations”. In: *Physical Review D* **79**.2, p. 024016.

Barceló, Carlos, Stefano Liberati, and Matt Visser (2005). “Analogue Gravity”. In: *Living Rev. Rel* **8**.12.

Barrett, Thomas William (2015). “On the Structure of Classical Mechanics”. In: *The British Journal for the Philosophy of Science* 66.4, pp. 801–828.

Bartha, Paul (2013). “Analogy and Analogical Reasoning”. In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Fall 2013.

Bartha, Paul FA (2010). *By Parallel Reasoning: The Construction and Evaluation of Analogical Arguments*. Oxford: Oxford University Press.

Beisbart, Claus and John D Norton (2012). “Why Monte Carlo Simulations Are Inferences and Not Experiments”. In: *International Studies in the Philosophy of Science* **26**.4, pp. 403–22.

Bekaert, Xavier, Nicolas Boulanger, and Per A Sundell (2012). “How Higher-spin Gravity Surpasses the Spin-two Barrier”. In: *Reviews of Modern Physics* 84.3, p. 987.

Bell, John S. (1966). “On the Problem of Hidden Variables in Quantum Mechanics”. In: *Reviews of Modern Physics* 38.3, p. 447.

Belot, Gordon (2011). “Background-independence”. In: *General Relativity and Gravitation* 43.10, pp. 2865–2884.

Bernstein, Ethan and Umesh Vazirani (1993). “Quantum Complexity Theory”. In: *Proceedings of the twenty-fifth annual ACM symposium on Theory of computing*. ACM, pp. 11–20.

Bjorken, James D (2001). “The Future of Particle Physics”. In: *International Journal of Modern Physics A* 16.04, pp. 483–502.

Bovens, Luc and Stephan Hartmann (2004). *Bayesian Epistemology*. Oxford: Oxford University Press.

Bub, Jeffrey (2010). “Von Neumann’s ‘No Hidden Variables’ Proof: a Reappraisal”. In: *Foundations of Physics* 40.9-10, pp. 1333–1340.

Callender, Craig (2005). “Answers in Search of a Question:‘Proofs’ of the Tri-Dimensionality of Space”. In: *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics* **36**.1, pp. 113–36.

Cartwright, Nancy (1999). *The Dappled World: A Study of the Boundaries of Science*. Cambridge: Cambridge University Press.

Carusotto, Iacopo et al. (2008). “Numerical Observation of Hawking Radiation from Acoustic Black Holes in Atomic Bose–Einstein Condensates”. In: *New Journal of Physics* **10**.10, p. 103001.

Cirac, J Ignacio and Peter Zoller (2012). “Goals and Opportunities in Quantum Simulation”. In: *Nature Physics* **8**.4, pp. 264–6.

Coleman, Sidney (1965). "Trouble with relativistic SU (6)". In: *Physical Review* 138.5B, B1262.

Coleman, Sidney and Jeffrey Mandula (1967). "All possible symmetries of the S matrix". In: *Physical Review* 159.5, p. 1251.

Corley, Steven (1998). "Computing the Spectrum of Black Hole Radiation in the Presence of High Frequency Dispersion: An Analytical Approach". In: *Physical Review D* **57**.10, p. 6280.

Cornwell, JF (1971). "On the Relevance of the Mass-Splitting Theorems". In: *Progress of theoretical physics* 45.6, pp. 1987–1988.

Corwin, L, Yu Ne'eman, and S Sternberg (1975). "Graded Lie Algebras in Mathematics and Physics (Bose-Fermi Symmetry)". In: *Reviews of Modern Physics* 47.3, p. 573.

Coutant, Antonin, Renaud Parentani, and Stefano Finazzi (2012). "Black Hole Radiation With Short Distance Dispersion, an Analytical S-Matrix Approach". In: *Physical Review D* **85**.2, p. 024021.

Crupi, Vincenzo (2013). "Confirmation". In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Winter 2013.

Cuffaro, Michael E (2016a). "On the Significance of the Gottesman–Knill Theorem". In: *The British Journal for the Philosophy of Science (forthcoming)*.

— (2016b). "Reconsidering No-Go Theorems from a Practical Perspective". In: *British Journal for the Philosophy of Science (forthcoming)*.

Curiel, Erik (2014). "Classical Mechanics is Lagrangian; it is Not Hamiltonian". In: *The British Journal for the Philosophy of Science* 65.2, pp. 269–321.

Dardashti, R et al. "Confirmation via Analogue Simulation: A Bayesian Account". (in progress).

Dardashti, Radin, Karim PY Thébault, and Eric Winsberg (forth.). "Confirmation via Analogue Simulation: What Dumb Holes Could Tell us About Gravity". In: *The British Journal for the Philosophy of Science*.

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

Dawid, Richard, Stephan Hartmann, and Jan Sprenger (2015). "The No Alternatives Argument". In: *The British Journal for the Philosophy of Science* 66.1, pp. 213–234.

Deutsch, David (1985). "'Quantum theory, the Church-Turing principle and the universal quantum computer'". In: *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences* **400**.1818, pp. 97–117.

Di Stefano, R (2000). "Notes on the Conceptual Development of Supersymmetry". In: *The Supersymmetric World: The Beginnings of the Theory*. Ed. by Gordan Kane and Mikhail A. Shifman. World Scientific, pp. 169–271.

Dirac, Paul A. M. (1940). "The Relation Between Mathematics and Physics." In: *Proceedings of the Royal Society of Edinburgh* 59, pp. 122–129.

— (1963). "The Evolution of the Physicist's Picture of Nature". In: *Scientific American* 208, pp. 45–53.

— (1980). "Why We Believe in the Einstein Theory". In: *Symmetries in Science*. Ed. by Bruno Gruber and Richard S. Millman. New York, London: Plenum Press, pp. 1–11.

Dizadji-Bahmani, Foad, Roman Frigg, and Stephan Hartmann (2011). "'Confirmation and Reduction: a Bayesian Account". In: *Synthese* **179**.2, pp. 321–38.

Donoghue, John F (1994). "General Relativity as an Effective Field Theory: The Leading Quantum Corrections". In: *Physical Review D* 50.6, p. 3874.

Einstein, Albert, Boris Podolsky, and Nathan Rosen (1935). "Can Quantum-Mechanical Description of Physical Reality be Considered Complete?" In: *Physical Review* 47.10, p. 777.

Ellis, George FR and Jean-Philippe Uzan (2005). "c is the Speed of Light, Isn't It?" In: *American journal of physics* 73.3, pp. 240–247.

Faulkner, Thomas et al. (2010). "From Black Holes to Strange Metals". In: *arXiv preprint arXiv:1003.1728*.

Feynman, Richard P. (1982). "Simulating Physics With Computers". In: *International Journal of Theoretical Physics* **21**.6, pp. 467–88.

— (1986). "Quantum Mechanical Computers". In: *Foundations of Physics* **16**.6, pp. 507–31.

Fine, Arthur (1982). "Joint Distributions, Quantum Correlations, and Commuting Observables". In: *Journal of Mathematical Physics* 23.7, pp. 1306–1310.

Fischer, Uwe R. and Ralf Schützhold (2004). "Quantum Simulation of Cosmic Inflation in Two-Component Bose-Einstein Condensates". In: *Physical Review A* **70**.6, p. 063615.

Garay, LJ et al. (2000). "Sonic Analog of Gravitational Black Holes in Bose-Einstein Condensates". In: *Physical Review Letters* **85**.22, p. 4643.

Gell-Mann, Murray (1964). "A Schematic Model of Baryons and Mesons". In: *Physics Letters* 8.3, pp. 214–215.

Georgescu, IM, Sahel Ashhab, and Franco Nori (2014). "Quantum Simulation". In: *Reviews of Modern Physics* 86.1, p. 153.

Georgi, Howard (1999). *Lie Algebras in Particle Physics. From Isospin to Unified Theories*. Vol. 54. Frontiers in Physics. Westview Press, Boulder USA.

Georgi, Howard and Sheldon L. Glashow (1974). "Unity of All Elementary-Particle Forces". In: *Physical Review Letters* 32.8, p. 438.

Gerritsma, Rene et al. (2010). "'Quantum simulation of the Dirac equation'". In: *Nature* **463**.7277, pp. 68–71.

Golfand, Yu. A. and E. P. Likhtman (1971). "Extension of the Algebra of Poincare Group Generators and Violation of p Invariance". In: *JETP Lett.* 13. [Pisma Zh. Eksp. Teor. Fiz.13,452(1971)], pp. 323–326.

Golfand, Yu A and EP Likhtman (1972). "On the Extensions of the Algebra of the Generators of the Poincaré Group by the Bispinor Generators". In: *I.E. Tamm Memorial Volume Problems of Theoretical Physics*. Ed. by V.L. Ginzburg et al. Moscow: Nauka, p. 37.

Greiner, Markus et al. (2002). "'Quantum phase transition from a superfluid to a Mott insulator in a gas of ultracold atoms'". In: *Nature* **415**.6867, pp. 39–44.

Griffiths, David (2008). *Introduction to Elementary Particles*. New York: John Wiley & Sons.

Gürsey, F., A. Pais, and L.A. Radicati (1964). "Spin and Unitary Spin Independence of Strong Interactions". In: *Physical Review Letters* 13.8, p. 299.

Harper, William L (2011). *Isaac Newton's Scientific Method: Turning Data into Evidence About Gravity and Cosmology*. Oxford: Oxford University Press.

Hartmann, Stephan (2015). "Imprecise Probabilities in Quantum Mechanics". In: *Foundations and Methods from Mathematics to Neuroscience.*

Essays Inspired by Patrick Suppes. Ed. by E. Colleen et al. Stanford: CSLI Publications.

Hartmann, Stephan and Jan Sprenger (2010). “Bayesian Epistemology”. In: *Routledge Companion to Epistemology,* ed. by Sven Bernecker and Duncan Pritchard. London: Routledge, pp. 609–620.

Hartnoll, Sean A, Christopher P Herzog, and Gary T Horowitz (2008). “Building a Holographic Superconductor”. In: *Physical Review Letters* 101.3, p. 031601.

Hawking, Stephen W. (1975). “Particle Creation by Black Holes”. In: *Communications in Mathematical Physics* **43**.3, pp. 199–220.

Healey, Richard (2005). *Gauging What’s Real: The Conceptual Foundations of Gauge Theories*. Oxford: Oxford University Press.

Heisenberg, W (1932). “Über den Bau der Atomkerne. I”. In: *Z. Phys* 77.1.

Hempel, Carl G (1965). *Aspects of Scientific Explanation and other Essays in the Philosophy of Science*. New York: Free Press.

Herzberg, Frederik (2014). “A Note on "The No Alternatives Argument" by Richard Dawid, Stephan Hartmann and Jan Sprenger”. In: *European Journal for Philosophy of Science* 4.3, pp. 375–384.

Hesse, Mary (1964). “‘Analogy and Confirmation Theory’”. In: *Philosophy of Science* **31**, pp. 319–27.

Hesse, Mary B (1966). *Models and Analogies in Science*. Vol. 7. Notre Dame: University of Notre Dame Press.

Himemoto, Yoshiaki and Takahiro Tanaka (2000). “‘Generalization of the model of Hawking radiation with modified high frequency dispersion relation’”. In: *Physical Review D* **61**.6, p. 064004.

Iorio, A (2011). “Alternative Symmetries in Quantum Field Theory and Gravity”. Charles University Habilitation (Prague).

Jacobson, Ted (1991). "Black-Hole Evaporation and Ultrashort Distances". In: *Physical Review D* **44**.6, p. 1731.

— (1993). "Black Hole Radiation in the Presence of a Short Distance Cutoff". In: *Physical Review D* **48**.2, pp. 728–41.

— (1996). "Introductory Lectures on Black Hole Thermodynamics". <www.physics.umd.edu/grt/taj/776b/lectures.pdf>.

Jauch, Josef Maria and Constantin Piron (1963). "Can Hidden Variables be Excluded in Quantum Mechanics". In: *Helv. Phys. Acta* 36, pp. 827–837.

Kalka, Harald and Gerhard Soff (1997). *Supersymmetrie*. Teubner Studienbücher Physik. Wiesbaden: Vieweg+ Teubner Verlag.

Kaluza, Theodor (1921). "Zum Unitätsproblem der Physik". In: *Sitzungsber. Preuss. Akad. Wiss. Berlin.(Math. Phys.)* 966972, p. 45.

Kaplan, David B (1992). "A Method for Simulating Chiral Fermions on the Lattice". In: *Physics Letters B* 288.3, pp. 342–347.

Keynes, John Maynard (1921). *A Treatise on Probability*. London: Macmillan & Co.

Kuipers, Theo AF (2002). "Beauty, a Road to the Truth". In: *Synthese* 131.3, pp. 291–328.

Landau, LD and EM Lifshitz (1987). *Fluid Mechanics*. Oxford: Butterworth-Heinemann, pp. 227–9.

Laudan, Larry (1978). *Progress and Its Problem: Towards a Theory of Scientific Growth*. London: University of California Press.

— (1996). *Beyond Positivism and Relativism*. Colorado: Westview Press.

Laudisa, Federico (2014). "Against the 'No-go' Philosophy of Quantum Mechanics". In: *European Journal for Philosophy of Science* 4.1, pp. 1–17.

Leonhardt, Ulf and Thomas G Philbin (2008). "The Case for Artificial Black Holes". In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **366**.1877, pp. 2851–7.

Liberati, Stefano (2013). "Tests of Lorentz Invariance: a 2013 Update". In: *Classical and Quantum Gravity* 30.13, p. 133001.

Lloyd, Seth (1996). "'Universal Quantum Simulators'". In: *Science* **273**, pp. 1073–7.

Ma, Xiao-song et al. (2011). "Quantum Simulation of the Wavefunction to Probe Frustrated Heisenberg Spin Systems". In: *Nature Physics* **7**.5, pp. 399–405.

Magueijo, Joao (2003). "New Varying Speed of Light Theories". In: *Reports on Progress in Physics* 66.11, p. 2025.

Mangano, Michelangelo L (2007). "The Future of Particle Physics". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 572.1, pp. 1–7.

Manousakis, Efstratios (2002). "A Quantum-Dot Array as Model for Copper-Oxide Superconductors: A Dedicated Quantum Simulator for the Many-Fermion Problem". In: *Journal of Low Temperature Physics* **126**.5-6, pp. 1501–13.

Mattingly, David (2005). "Modern Tests of Lorentz Invariance". In: *Living Rev. Rel* 8.5, p. 2003.

McAllister, James W. (1990). "Dirac and the Aesthetic Evaluation of Theories". In: *Methodology and Science* 23, p. 87.

— (1999). *Beauty and Revolution in Science*. Ithaca and London: Cornell University Press.

McGlinn, W.D. (1964). "Problem of Combining Interaction Symmetries and Relativistic Invariance". In: *Phys.Rev.Lett.* 12, pp. 467–469.

Mermin, N. David (1993). "Hidden Variables and the Two Theorems of John Bell". In: *Reviews of Modern Physics* 65.3, p. 803.

Michel, Louis (1965). "Relations Between Internal Symmetry and Relativistic Invariance". In: *Physical Review* 137.2B, B405.

Mill, John Stuart (1843). *A System of Logic, Ratiocinative and Inductive: Being a Connected View of the Principles of Evidence and the Methods of Scientific Investigation*. Vol. 7 & 8. Toronto: University of Toronto Press.

Mirman, R. (1969). "The Physical Basis of Combined Symmetry Theories". In: *Progress of Theoretical Physics* 41.6, pp. 1578–1584.

Mueller-Kirsten, H. and Armin Wiedemann (1987). *Supersymmetry - An Introduction with Conceptual and Calculational Details*. Teaneck, NJ: World Scientific Pub. Co.

Mukhanov, Viatcheslav and Sergei Winitzki (2007). *Introduction to Quantum Effects in Gravity*. Cambridge: Cambridge University Press.

Ne'eman, Yuval (1961). "Derivation of Strong Interactions From a Gauge Invariance". In: *Nuclear physics* 26.2, pp. 222–229.

Ne'eman, Yuval and Shlomo Sternberg (1991). "Internal Supersymmetry and Superconnections". In: *Symplectic Geometry and Mathematical Physics*, p. 326.

Neumann, Johann (1932). *Mathematische Grundlagen der Quantenmechanik*. Berlin: Springer.

Nielsen, Holger Bech and Masao Ninomiya (1981). "Absence of Neutrinos on a Lattice:(I). Proof by Homotopy Theory". In: *Nuclear Physics B* 185.1, pp. 20–40.

Nielsen, Michael A and Isaac L Chuang (2010). *Quantum computation and quantum information*. Cambridge: Cambridge University Press.

Nijenhuis, Albert (1955). "Jacobi-Type Identities for Bilinear Differential Concomitants of Certain Tensor Fields." In: *Indagationes Mathematicae (Proceedings)*. Vol. 58. Elsevier, pp. 390–397.

Noh, Changsuk, BM Rodríguez-Lara, and Dimitris G Angelakis (2012). "Quantum Simulation of Neutrino Oscillations with Trapped Ions". In: *New Journal of Physics* **14**.3, p. 033028.

Nola, Robert and Howard Sankey (2014). *Theories of Scientific Method: an Introduction*. New York: Routledge.

Nordström, Gunnar (1914). "Über die Möglichkeit, das Elektromagnetische Feld und das Gravitationsfeld zu Vereinigen". In: *Physikalische Zeitschrift* 15, p. 504.

North, Jill (2009). "The "Structure" of Physics: A Case Study". In: *The Journal of Philosophy* 106.2, pp. 57–88.

Novello, Mário, Matt Visser, and Grigory E Volovik (2002). *Artificial Black Holes*. Singapore: World Scientific Singapore.

O'Raifeartaigh, Lochlainn (1965). "Lorentz Invariance and Internal Symmetry". In: *Physical Review* 139.4B, B1052.

Parker, Wendy S (2009). "Does Matter Really Matter? Computer Simulations, Experiments, and Materiality". In: *Synthese* **169**.3, pp. 483–96.

Pati, Jogesh C and Abdus Salam (1974). "Lepton Number as the Fourth "Color"". In: *Physical Review D* 10.1, p. 275.

Pitowsky, Itamar (1989). *Quantum Probability, Quantum Logic*. Vol. 321. Lecture Notes in Physics. Berlin, Heidelberg, New York: Springer.

Pooley, Oliver (forthcoming). "Background Independence, Diffeomorphism Invariance, and the Meaning of Coordinates". In: *Towards a Theory of Spacetime Theories*. Ed. by Dennis Lehmkuhl, Gregor Schiemann, and Erhard Scholz. Springer.

Popper, Karl (1989). *Conjectures and Refutations: The Growth of Scientific Knowledge*. London: Routledge.

Randall, Lisa and Raman Sundrum (1999). "Large Mass Hierarchy From a Small Extra Dimension". In: *Physical Review Letters* **83**.17, p. 3370.

Reid, Thomas and William Hamilton (1850). *Essays on the Intellectual Powers of Man*. Cambridge: J. Bartlett.

Rickles, Dean (2008). "Who's Afraid of Background Independence?" In: *Philosophy and Foundations of Physics* 4, pp. 133–152.

— (2014). *A Brief History of String Theory*. Berlin, Heidelberg: Springer.

Roberts, Bryan W (2011). "Group Structural Realism". In: *The British Journal for the Philosophy of Science* 62.1, pp. 47–69.

Russell, Neil and Alan Kostelecky (2009). "Data Tables for Lorentz and CPT Violation". In: *APS Division of Atomic, Molecular and Optical Physics Meeting Abstracts*. Vol. 1, p. 5008.

Sakita, B (1964). "Supermultiplets of Elementary Particles". In: *Physical Review* 136.6B, B1756.

Schrödinger, Erwin (1930). "Über die Kräftefreie Bewegung in der Relativistischen Quantenmechanik". In: *Sitz. Preuss. Akad. Wiss. Phys.-Math. Kl.* 24, pp. 418 –428.

Schützhold, Ralf and W.G. Unruh (2002). "Gravity Wave Analogues of Black Holes". In: *Physical Review D* **66**.4, p. 044019.

Shamir, Yigal (1993). "Chiral Fermions from Lattice Boundaries". In: *Nuclear Physics B* 406.1-2, pp. 90–106.

Shor, Peter W. (1997). "Polynomial-Time Algorithms for Prime Factorization and Discrete Logarithms on a Quantum Computer'". In: *SIAM Journal on Computing* **26**.5, pp. 1484–509.

Sklar, Lawrence (1975). "Methodological Conservatism". In: *The Philosophical Review* 84.3, pp. 374–400.

Slansky, Richard (1981). “Group Theory for Unified Model Building”. In: *Physics Reports* 79.1, pp. 1–128.

Smolin, Lee (2014). “String Theory and the Scientific Method”. In: *American Journal of Physics* 82.11, pp. 1105–1107.

Sohnius, Martin F. (1985). “Introducing Supersymmetry”. In: *Physics Reports* 128.2, pp. 39–204.

Stanford, Kyle (2016). “Underdetermination of Scientific Theory”. In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Spring 2016.

Stanford, P. Kyle (2001). “Refusing the Devil’s Bargain: What Kind of Underdetermination Should We Take Seriously?” In: *Philosophy of Science* 68.3, pp. S1–S12.

Stanford, P Kyle (2006). *Exceeding Our Grasp: Science, History, and the Problem of Unconceived Alternatives*. Oxford: Oxford University Press.

Sudarshan, ECG (1965). “Concerning Space-Time, Symmetry Groups, and Charge Conservation”. In: *Journal of Mathematical Physics* 6.8, pp. 1329–1331.

Suppes, Patrick and Mario Zanotti (1991). “Existence of Hidden Variables Having Only Upper Probabilities”. In: *Foundations of Physics* 21.12, pp. 1479–1499.

Teller, Paul (2000). “The Gauge Argument”. In: *Philosophy of Science* 67.2, pp. 466–481.

Thompson, R.T. and L.H. Ford (2008). “Enhanced Black Hole Horizon Fluctuations”. In: *Physical Review D* **78**.2, p. 024014.

Timpson, Christopher G. (2013). *Quantum Information Theory & the Foundations of Quantum Mechanics*. Oxford: Oxford: Oxford University Press.

Toomre, A. and J. Toomre (1972). "Galactic Bridges and Tails". In: *Astrophysical Journal* 178, pp. 623–66.

Unruh, W.G. (1981). "Experimental Black-Hole Evaporation?" In: *Physical Review Letters* **46**.21, pp. 1351–53.

— (1995). "Sonic analogue of Black Holes and the Effects of High Frequencies on Black Hole Evaporation'". In: *Physical Review D* **51**.6, p. 2827.

— (2008). "Dumb Holes: Analogues for Black Holes". In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **366**.1877, pp. 2905–13.

— (2014). "Has Hawking Radiation Been Measured?" In: *Foundations of Physics* 44.5, pp. 532–545.

Unruh, W.G. and Ralf Schützhold (2005). "Universality of the Hawking Effect". In: *Physical Review D* **71**.2, p. 024028.

Volkov, D. V. and V. P. Akulov (1972). "Possible Universal Neutrino Interaction". In: *JETP Lett.* 16. [Pisma Zh. Eksp. Teor. Fiz.16,621(1972)], pp. 438–440.

Weinberg, S (2009). "Effective Field Theories, Past and Future". In: *6th International Workshop on Chiral Dynamics*. Vol. 1, p. 1.

Weinberg, Steven (1996). *The Quantum Theory of Fields: Modern Applications*. Vol. 2. Cambridge: Cambridge University press.

— (2011). *The Quantum Theory of Fields: Supersymmetry*. Vol. 3. Cambridge: Cambridge University Press.

Weinberg, Steven and Edward Witten (1980). "Limits on Massless Particles". In: *Physics Letters B* 96.1, pp. 59–62.

Weinfurtner, Silke et al. (2011). "Measurement of Stimulated Hawking Emission in an Analogue System". In: *Physical Review Letters* **106**.2, p. 021302.

Weinfurtner, Silke et al. (2013). "Classical Aspects of Hawking Radiation Verified in Analogue Gravity Experiment". In: *Analogue Gravity Phenomenology*. Ed. by Daniele Faccio et al. Heidelberg: Springer, pp. 167–180.

Wess, Julius and Bruno Zumino (1974a). "A Lagrangian Model Invariant under Supergauge Transformations". In: *Physics Letters B* 49.1, pp. 52–54.

— (1974b). "Supergauge Transformations in Four Dimensions". In: *Nuclear Physics B* 70.1, pp. 39–50.

Wilson, Fred (2016). "John Stuart Mill". In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Spring 2016.

Winsberg, Eric (2009). "A Tale of Two Methods". In: *Synthese* **169**.3, pp. 575–92.

— (2010). *Science in the Age of Computer Simulation*. Chicago: University of Chicago Press.

Zee, Anthony (2010). *Quantum Field Theory in a Nutshell*. 2nd. Princeton: Princeton University Press.