This electronic thesis or dissertation has been downloaded from the University of Bristol Research Portal, http://research-information.bristol.ac.uk

Author:
Antoniou, Antonis

Title:
Models, data, and unobservable phenomena in physics

General rights
Access to the thesis is subject to the Creative Commons Attribution - NonCommercial-No Derivatives 4.0 International Public License. A copy of this may be found at https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode This license sets out your rights and the restrictions that apply to your access to the thesis so it is important you read this before proceeding.

Take down policy
Some pages of this thesis may have been removed for copyright restrictions prior to having it been deposited on the University of Bristol Research Portal. However, if you have discovered material within the thesis that you consider to be unlawful e.g. breaches of copyright (either yours or that of a third party) or any other law, including but not limited to those relating to patent, trademark, confidentiality, data protection, obscenity, defamation, libel, then please contact collections-metadata@bristol.ac.uk and include the following information in your message:

•Your contact details
•Bibliographic details for the item, including a URL
•An outline nature of the complaint

Your claim will be investigated and, where appropriate, the item in question will be removed from public view as soon as possible.

Models, Data, and Unobservable Phenomena in Physics

ANTONIS ANTONIOU

University of BRISTOL

DEPARTMENT OF PHILOSOPHY

A dissertation submitted to the University of Bristol
in accordance with the requirements for award of the
degree of Doctor of Philosophy in the Faculty of Arts.

APRIL 2022

Word count: 75,674

ABSTRACT

This thesis provides a model-based philosophical investigation of the epistemology and methodology of modern physics, based on three main axes: (i) the relationship between models and background theories, (ii) the relationship between models and experimental data, and (iii) the observation of unobservable entities in physics, such as dark matter particles.

The first part of the thesis (Chapters 1 and 2) is more generic in nature. It comprises an analysis and synthesis of the literature on the structure of scientific theories and its chronological development into a literature on the nature and role of models in science. It also provides a chapter-length critical evaluation of the literature on the ontology of models accompanied with a novel Carnapian solution to the metaphysical challenges that appear in the debate.

In the second part of the thesis (Chapters 3, 4, and 5) a more practical approach is adopted, and the three main axes of the thesis are developed respectively. Chapter 3 concerns the relationship of models with background theories in perturbative quantum field theory, in which a new type of models that does not belong to the traditional dichotomy between theoretical and phenomenological models is identified. Chapter 4 concerns the relationship of models with experimental data via the construction of data models. A detailed case study of the experimental tests of the Standard Model at the LHCb experiment is conducted, which leads to a number of interesting conclusions about the nature of data models, and the two distinctions between raw/processed data and real/simulated data. Finally, Chapter 5 provides a novel framework for the epistemology of dark matter observation, via which it is argued that a partial explanation for the slow progress in this field is due to the fact that robustness arguments from the variability of experiments are significantly limited.

DEDICATION AND ACKNOWLEDGEMENTS

The successful completion of this thesis would not have been possible without the valuable contribution of several people who kindly offered their support and expertise during my doctoral studies at the University of Bristol.

Firstly, I would like to thank my friend and philosophical mentor Demetris Portides. His lectures on various topics of philosophy of science and his overall approach to the intersection of science and philosophy inspired me to pursue postgraduate studies during my time as an undergraduate student in Philosophy at the University of Cyprus. Demetris' support, encouragement and guidance during the final stages of my undergraduate studies were decisive for the pursuit of postgraduate studies at the University of Bristol, and for this I am deeply grateful to him. I would also like to thank Demetris for reading parts of this thesis and proposing substantial improvements, especially in Chapter 2. I am also grateful to Vassilis Livanios and Christos Kyriacou from the University of Cyprus for their valuable support and guidance throughout different stages of my philosophical journey during the last years.

Secondly, I would like to thank a team of great philosophers at the University of Bristol and beyond for kindly reading parts of this thesis and providing valuable feedback, and for countless stimulating philosophical discussions that inspired parts of this work. These are (in alphabetical order): Luca Castaldo, David Cobb, Ana-Maria Crețu, Adam Estrup, Alex Franklin, Steven French, Lucy James, James Ladyman, Laurie Letertre, Samir Okasha, Vanessa Seifert, Nicos Stylianou, and Lena Zuchowski. To Alex Franklin, I am especially grateful for his valuable insights into perturbative quantum field theory during a series of a mini reading group between me, him and my supervisor Karim Thébault, which led to the successful completion of Chapter 3. Many thanks also to James Fraser and Michael Miller for our discussions on their work on perturbative quantum field theory which made the completion of the relevant chapter possible, as well as to four anonymous referees for their valuable reviews of Chapters 2 and 4 as journal articles.

I am also greatly appreciative to Jonas Rademacker and Konstantinos Petrides from the Bristol Flavour Physics Group for their valuable insights on various aspects of the LHCb experiment at CERN and for their guidance through the literature on Lepton Flavour Universality tests, which led to the succesful completion of Chapter 4.

A special thanks also goes to my secondary supervisor Adam Toon at the University of Exeter. Adam has kindly read my writings several times and provided valuable and substantial feedback in every chapter. Our conversations during my visits in Exeter

were particularly illuminating and helpful, especially during my first steps into the vast literature on scientific models.

Last and foremost, I would like to express my deepest gratitude and appreciation to my supervisor Karim Thébault. Karim has been a great supervisor and friend from the first day of my doctoral studies at the University of Bristol and to him I owe my development from a simple fan of philosophy and science, to a researcher in philosophy of science. The final outcome of this thesis is a result of Karim's constant guidance and encouragement to develop and refine my ideas.

This work was generously funded by the Arts and Humanities Reasearch Council (AHRC) via the South, West and Wales Doctoral Training Partnership (SWW DTP), Award Reference: 2109811.

AUTHOR'S DECLARATION

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's Regulations and Code of Practice for Research Degree Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

SIGNED: ANTONIS ANTONIOU DATE: 20 APRIL 2022

TABLE OF CONTENTS

LIST OF FIGURES

List of Tables

PREFACE

Ever since I can remember myself as a high-school student, I have always had a deep interest in physics and scientific methodology in general. And although I never had the chance to formally study physics at a higher level, this interest was kept alive until my late days as a mature undergraduate student in Philosophy at the University of Cyprus back in 2017. It was during the last year of my studies when I had the chance to take a course on philosophy of science and realise that the two academic subjects that intrigue me the most – physics and philosophy – can be studied together. During these lectures I learned a great deal about various core topics in philosophy of science, but I particularly remember a lecture on scientific models.

The lecturer draw an inclined plane with a sliding body on the board and started asking several questions regarding the idealisations of the model. 'We all know that in the real world there is friction between the sliding body and the surface of the plane, but we still chose to remove it from our calculations' he said. 'We also know that the mass distribution of real objects is spread throughout the objects, but we instead represent them in our calculations as point mass objects. How is it possible then to use such simple models to learn so many things about complicated real physical systems and make accurate predictions for their behaviour, given that they are – in a sense – full of mistakes?'. I was genuinely puzzled with this question. For the first time, I had realised that my limited knowledge of Newtonian physics that I acquired from high-school was almost entirely based on a host of assumptions on which these simple models were built. But at the same time, I realised how powerful these models are. In all their simplicity, they are valuable epistemological tools that offer – above anything else – the necessary mathematical tractability for studying the behaviour of messy, complicated and multifarious physical systems.

A few months later, I found myself in a Master's programme at the University of Bristol, trying to write a research proposal for my upcoming doctoral studies. The philosophical investigation of the role of models in physics seemed to be a very interesting and rich topic to explore, and so I started reading more about this topic. I then realised that I was not alone in trying to understand how exactly models work and what makes them such exceptional epistemological tools. My first contact with the vast literature on scientific models brought to my attention the endless possibilities of a philosophical investigation on the nature and the role of models in science. Indeed, the present thesis was initially inspired by the literature on scientific representation with models and soon developed into a model-based philosophical investigation of the epistemology and methodology of modern physics. The main objective of this project is to study the methodology of modern physics based on three major axes: (i) the relationship between models and physical theories, (ii) the relationship between models and experimental data, and (iii) the observation of unobservable entities in physics, such as dark matter particles.

In most parts, especially in Chapters 3, 4 and 5, the thesis maintains a practice-oriented spirit by closely examining examples of real scientific practice in order to answer various philosophical questions. This approach is nicely outlined by Ankeny et al. (2011) in an introductory paper of a special issue of the *European Journal for Philosophy of Science* on ‘Philosophy of Science in Practice’. According to the authors, the aim of Philosophy of Science in Practice is to foster ‘the pursuit of a philosophy of science that considers theory, practice and the world simultaneously, and never in isolation from each other’ (*ibid.*, p.304). Scientific practice is understood in this context as the sum of organised and regulated activities by scientists aiming at the achievement of certain goals. These goals might be the exploration of new untrodden ground in experiments, the testing of a specific theoretical hypothesis, the application of new scientific findings and so on. By doing Philosophy of Science in Practice, the aim is to examine traditional philosophical questions regarding epistemological concepts such as truth, belief, observation, justification, explanation etc. vis-à-vis the actual scientific practice, by taking into account the implications of the practical nature of the sciences on these issues.

Of course, as aptly noted by an anonymous referee during the reviewing process of Chapter 4 as a journal article, in doing Philosophy of Science in Practice there always looms the danger of engaging in a mere description of scientific practice, neglecting the importance of connecting the practical aspect of science with the underlying philosophical questions. I must confess that I was perhaps guilty of doing so in earlier drafts of some chapters, especially in Chapter 4, in which I initially engaged in a detailed description of the various aspects of the LHCb experiment at CERN and the theoretical foundations of Lepton Flavour Universality tests. In all fairness, most of these descriptions were the result of my personal attempt to learn and understand the necessary physics behind the philosophical questions I was after, and writing them down in a clear and concise manner was extremely beneficial. Nevertheless, what I have learned during the revisions of the chapters is that the most important thing in doing Philosophy of Science in Practice is to find the right balance between the philosophical discussions and the exposition of scientific facts and details in order to produce a well-balanced and scientifically informed philosophical study. Hopefully, I have managed to do so in this thesis.

The thesis consists of six chapters. Chapter 1 is introductory in nature. It begins with a historical review of the literature on the structure of scientific theories and its gradual transformation into the vast literature on scientific models. The discussion in this chapter includes the presentation of a novel conceptual framework for the description of models in physics in terms of their four most characteristic features (a mathematical framework, an interpretation, an ideal system, and the representational media), followed by a discussion of a very important distinction in the methodology of physics, namely, the distinction between theoretical and phenomenological models. The chapter ends with a review of the literature regarding the three major aims of physics (economic description, explanation, and prediction) and the role of models in achieving these goals.

Chapter 2 is a philosophical intervention in the long-standing debate on the ontology of models. My first contact with the literature on the ontology of models made me realise that a large part of the debate concerns various traditional metaphysical difficulties regarding the nature of abstract objects and the attribution of properties to them. Drawing on Carnap's approach to metaphysics, it is argued in this chapter that many of the difficulties that essentially hinder the resolution of the debate stem from an inappropriate reading of the question of the ontology of models as a purely metaphysical question. The suggestion is to instead view this question as either (i) an internal theoretical question within an already accepted linguistic framework or (ii) an external practical question regarding the choice of the most appropriate form of language in order to describe and explain the practice of scientific modelling. The main implication of this view is that the question of the ontology of models becomes a means of probing other related questions regarding the overall practice of scientific modelling, such as questions on the capacity of models to provide knowledge, the relation of models with background theories etc., with which the remaining of the thesis is concerned.

Chapter 3 marks the beginning of the Philosophy of Science in Practice approach in the thesis. By studying the modelling techniques of perturbative quantum field theory and their reliance on regularisation and renormalisation techniques, it is argued that models of perturbative quantum field theory are not theoretical, nor phenomenological. Rather, they comprise a special third type of models in physics, in that they are products of non controllable idealisations. As such, they pose some challenges to the semantic view of theories and to van Fraassen's claim that along with the observable phenomena, the theoretical models of a physical theory are the poles of scientific understanding. It is also argued that the identification of this third type of models in physics forces us to reconsider the relationship between theories and models, as well as the empirical adequacy of the former in terms of the successful predictions of the latter.

Chapter 4 is perhaps the most characteristic example of Philosophy of Science in Practice in the thesis. Drawing on Suppes' well-known claim that theoretical hypotheses are confronted with data models rather than the raw data of an experiment, a detailed case study of Lepton Flavour Universality tests at the LHCb experiment is presented, in order to understand the nature and the role of data models in high-energy physics. The close examination of the scientific practice at the LHCb provides a solid understanding of what Suppes had in mind when referring to data models as 'simple entities' which, nonetheless, emerge from the 'maddeningly diverse and complex experience' we call experiment. Throughout the chapter, the importance of considerations regarding the selection criteria, efficiency calculations, data fitting, and uncertainties in the process of constructing a data model are highlighted, and some interesting conclusions regarding the two distinctions between raw/processed data and real/simulated data are extracted.

Chapter 5 concerns the scientific methodology for the observation of dark matter in astrophysics. In this chapter, a novel framework for the epistemology of dark matter observation is provided via a conceptual analysis of the five main methods of dark matter observation. Based on the epistemic virtues of these five methods with respect to their informativeness, model sensitivity and reliability, it is argued that a partial explanation for the slow progress in dark matter research comes from the important limitations of robustness arguments from the variability of experiments in this field. The discussion closes with what we shall call 'the puzzle of dark matter observation', namely, the fact that the only possible way to collect reliable results that will eventually narrow down the range of viable phenomenological models for dark matter, is by presupposing these models at the first place.

Finally, Chapter 6 concludes the thesis by bringing together the core ideas of each chapter and highlighting the key conclusions of the thesis. An attempt is made to recapitulate the main ideas of Chapter 1 with respect to the four main characteristics of models, and elucidate the ways in which this conceptual framework facilitates the discussions in the following chapters. The chapter also includes the presentation of a general framework of the methodology of modern physics as it was outlined throughout the thesis, and closes with a discussion of the most important open questions that emerge from the present study, pointing out some possible directions for future work.

Statement of Publications By Author

Chapter 2: A slightly shorter version of this chapter, excluding Section 2.2.3, has been published as:

- Antoniou, A. (2021). A pragmatic approach to the ontology of models. *Synthese 199*(3). 6645-6664. https://doi.org/10.1007/s11229-021-03085-9

Chapter 4: A slightly shorter version of this chapter, mainly excluding some technical details from Section 4.2, has been published as:

- Antoniou, A (2021) What is a data model? *European Journal for Philosophy of Science, 11*(4), 1-33. https://doi.org/10.1007/s13194-021-00412-2

CHAPTER 1

INTRODUCTION: MODELS AND THE AIMS OF PHYSICS

The central theme of this thesis is the interplay between theories, models, and data from unobservable phenomena in physics. It could be said that these three elements form the backbone of the methodology of modern physics: we observe nature by collecting data, we represent its countless manifestations by constructing models, and then we fit these models in physical theories that aim to classify, explain, and predict physical phenomena. Among these three basic elements, models have, by far, attracted the most attention from philosophers during the last four decades. Indeed, the literature on scientific models is vast, and comprises topics ranging from their ontology and their representational capacity, to their function as explanatory and exploratory tools.

The detailed study of scientific models during the last decades has undoubtedly provided important insights on these issues. Compared to our understanding of the role of models in science back in the 1980's, one can safely argue that today we have made significant progress in understanding how models contribute to the acquirement of knowledge in almost every branch of science, from 'soft' sciences such as psychology, economics and social science, to the 'hard' natural sciences of biology, chemistry and physics. Nevertheless, while many general questions about scientific modelling have been answered to a greater or lesser extent, there are still several specific questions that have yet to be resolved.

The aim of this thesis is to shed light in some of these questions related to the role of models in the methodology of physics. These questions will be put into words in each one of the following chapters. Meanwhile, in this first chapter, we shall be concerned with an analysis of – parts of – the literature on scientific models, in order to lay the conceptual foundations for this thesis. In particular we will begin in Section 1.1 with a historical review on the literature on the structure of scientific theories, which as we shall see, is the predecessor of the vast philosophical literature on models. This historical perspective will illustrate how the different approaches to the structure of scientific theories resulted in different understandings of the concepts of models and their relationship with theories. In Section 1.2 we will proceed to the presentation of a novel conceptual framework about the nature of models in physics with an example from quantum electrodynamics (QED). This framework describes models in terms of their four main characteristics: (i) a mathematical framework, (ii) an interpretation (iii) an ideal system and (iv) the representational media by which it is expressed. As will be made clear, this characterisation of models does not amount to a definition, nor is it an attempt to provide the necessary and sufficient conditions for something to count as a model. Rather, it is merely a useful conceptual framework to facilitate our understanding of the nature of models in physics. The word 'models' will be used in this thesis perhaps more than any other word, and it is useful to have a clear conceptual understanding of what is meant by it. In the following section (Section 1.3), we shall be concerned with an important distinction which will be constantly present throughout this thesis and which plays a central role in the methodology of physics, namely the distinction between theoretical and phenomenological models. The upshot of the discussion in this section is that this distinction – at least insofar as it is defined in this thesis in these specific terms – does not reflect a difference in the intrinsic nature of theoretical and phenomenological models or their representational capacity. Rather, it is a distinction that concerns the main aims for which these models are built, as well as their role in the methodology of physics. Finally, we shall close this chapter with an evaluation of the contribution of models to the fulfilment of the broader aims of physics. That is, in Section 1.4, it will be argued that the three major aims (or achievements) of physics are (i) the economic description of nature, (ii) the explanation of physical phenomena, and (iii) the prediction of results. The discussion in this section will also comprise a brief analysis on each of these topics, inspired by the work of prominent historical figures in philosophy of science such as Pierre Duhem, Ernst Mach, Carl Hempel and Imre Lakatos, in an attempt to link some well known philosophical views of the past with more recent views on the role of models in modern physics.

Before we proceed however, it is imperative to clarify two important issues that the reader should bear in mind throughout the thesis. The first is that in what follows we shall be mainly concerned with *representational models*, i.e. with models that serve as representations of physical systems that are known to exist – e.g. a pair of electrons – or that are hypothetically existing – e.g. dark matter particles. What we shall not be concerned with are models whose target systems are *known to not exist*, such as models of the ether and models of three-sex populations. The latter are found in the literature under various names such as fictional models (Morrison 2015, Ch.3) and targetless models (Weisberg 2013), and should be distinguished from *toy models*, i.e. simple and highly idealised models which nonetheless act as representations – albeit abstract ones – of physical systems (Reutlinger et al. 2020; Nguyen 2020). Moreover, we shall endorse a pragmatic approach to representation as a triadic relation between the model, its users, and the represented target system. This view has been expressed in various ways in the literature (e.g. Bailer-Jones 2003; Suarez 2004; Giere 2004), and the common idea is that models represent their targets by virtue of the mental states of their users, and thus, there are no objective criteria that a model needs to fulfil in order to count as a representation of a physical system. In other words, anything can represent anything, insofar as the users/makers of a model engage in, what Callender and Cohen (2006, p.75), call an act of a *stipulative fiat* according to which a scientific model M represents a target system T insofar as its user stipulates that M represents T.[1]

The second thing to bear in mind is that models will be understood throughout this thesis as *multifunctional epistemological tools* for the acquirement of knowledge about the properties and the behaviour of the physical systems that they are supposed to represent. This view has its roots in the seminal work of Morgan and Morrison (1999) and has been endorsed, implicitly or explicitly, by the majority of philosophers. The most explicit formulation of this view is probably found in the works of Tarjia Knuuttila and Mieke Boon (Knuuttila 2005; 2021; Boon & Knuuttila 2009; 2011). The core idea in these works is that the epistemic value of scientific models does not principally come from their representational capacity as is often implied, but rather, from the practice of modelling as a whole. This practice includes the various ways in which models are conceived, constructed, manipulated and adjusted to the context specific purposes of scientific reasoning, theory construction and experimental design.

---

[1] The choice of studying mainly representational models concerns the fact that most of the issues to be examined in the thesis are based on the relationship of these models with the physical systems they represent and the ways in which they can provide knowledge about the existence or non-existence of their targets and their behaviour. However, this should not be taken as a statement that the sole purpose of scientific models is to represent physical systems and that non-representational models cannot provide knowledge about nature. For an appraisal of the role of non-representational models in science see Weisberg (2013, Ch.7) and Grune-Yanoff (2013).

The epistemic nature of models as tools of investigation is also evident in the discussions about the various functions of models. Redhead (1980), for instance, notes that the three main uses of models in physics are: (i) to probe parts of theories that may require further investigation, (ii) as heuristic tools in enriching theories to accommodate phenomena that lie outside their scope, and (iii) to test the predictions of theories. In a similar spirit, Gelfert (2016, pp.83-97) emphasises the *exploratory* function of models in science (i) as the starting points for future inquiries which may lead to more sophisticated versions of models, (ii) as tools for proof-of-principle demonstrations showcasing the feasibility of certain processes in nature, (iii) in providing potential explanations of why and how things happen, and (iv) for assessing the suitability of a target system under the scope of a background theory.

Drawing on this exploratory nature of models, Massimi (2018) highlights the presence of a special type of exploratory models in physics, which she calls *perspectival models*. The distinctive feature of perspectival models is their *sui generis* representational content which, contra to the traditional understanding of scientific representation, does not amount to a mapping relationship between the properties attributed to the models and the physical properties of their targets. Rather, it is characterised by a modal aspect, in that it is about 'exploring and ruling out the space of possibilities in domains that are still very much open-ended for scientific discovery' (*ibid.*, p.338). The primary function of perspectival models is thus exploratory: 'they are crucial tools for scientific discovery in designated areas of scientific inquiry, where methodological challenges about the search for new kinds of entities arise'(*ibid.*). This is achieved by the capacity of these models to provide modal knowledge of what might or might not be real within the context of a background theory and the results of an experiment. In accordance with Knuuttila and Boon, Massimi also notes that the exploratory function of perspectival models does not rely on their representational capacity, but rather on their unique ability to explore the space of possibilities for the behaviour and the properties of new kinds of entities.

These contemporary discussions on the epistemic value of scientific models have made the crucial role of models in the acquirement of knowledge within modern scientific practice more than salient. This view now transcends an older instrumentalist view on the role of models in science, according to which models were thought to serve merely an auxiliary function as computational devices for testing theories by generating their predictions, and for channelling the existential claims of the postulates of theories to some confrontation with experimentally testable consequences.[2] The historical review in

---

[2] This view is reported – but not endorsed– by Wartofsky (1979), and restated in Gelfert (2016,

the next section illustrates how this instrumentalist approach of models originated in the Syntactic View on the structure of scientific theories by the Logical Positivists, and slowly transformed into a general consensus about the crucial role of models in science as carriers of knowledge.

1.1 From theories to models

In the philosophical literature on the structure of scientific theories one finds three basic accounts: the Syntactic View, the Semantic View, and the Pragmatic View. The Syntactic View (also known as the Received view) is chronologically the first detailed account for the structure of scientific theories, developed in the heart of the Vienna circle from the 1920s to the 1960s by Logical Positivists, especially Hempel (1958) and Carnap (1966; 1967). According to this view, scientific theories are best described as axiomatised collections of sentences in higher order predicate logic. Roughly speaking, the logical framework of a scientific theory consists of *terms*, *sentences* and *languages* that can be further classified as *theoretical* or *observational.* Theoretical sentences consist of logical/mathematical symbols (i.e. quantifiers, connectives and other mathematical operators) and *theoretical terms.* The latter are constructs of the theory that concern concepts and properties relative to the theory such as 'atom', 'proton', 'temperature' etc. Likewise, observational sentences comprise logical/mathematical symbols and *observational terms*, that is, terms that refer to observable properties such as 'hot', 'hard', 'at position x' etc. Theoretical and observational terms are connected through a special set of sentences called *correspondence rules*, that contain both theoretical and observational terms, as well as logical/mathematical operations. For instance, Carnap (1966, p.233) gives the following example of a correspondence rule: 'The temperature of a gas is proportional to the mean kinetic energy of its molecules' in which a theoretical (non-observable) term in molecular theory is connected with an observable – the temperature of the gas. Correspondence rules thus provide the theoretical syntax of a theory with an interpretational framework and an application, i.e., a semantics, and they are responsible for linking theory with observation. In all, a scientific theory is understood by the Syntactic view as a syntactically formulated set of theoretical sentences (axioms, theorems, and natural laws) together with their interpretation via correspondence sentences.

The Syntactic View of scientific theories was met with heavy criticism in later years, mainly due to its strong dependence on the formalisation of predicate logic. In an

---

p.39).

overview on the fall of the Received View, Suppe (2000, p.S103) identifies two major problems. The first is that scientific theories are not linguistic entities as the Syntactic View holds, and thus the view individuates theories based on the wrong criterion. The second and most important, is that the correspondence rules taken by the Syntactic View as a necessary and important part of theories, are in fact a heterogeneous confusion of meaning relationships, experimental design, measurement, and causal relationships, some of which are not properly parts of theories. These two problems, along with a gradually growing belief among the philosophical community that symbolic logic is not the most appropriate formalism for the reconstruction of the structure of scientific theories led to the development of the Semantic View.

The overarching theme of the Semantic view of scientific theories is that theory structure is best captured by the language and tools of mathematics rather than predicate logic. In this view, scientific theories are understood as collections of mathematical models that are derivable from the fundamental axioms of the theory. Within the Semantic View one finds two main approaches: (i) the *set-theoretic approach* inspired by the work of Alfred Tarski on model theory and further developed by Patrick Suppes and his collaborators (1957; 1967; 2002), and (ii) the *state-space approach* initiated by the works of Evert Beth and Herman Weyl and later articulated in detail by Bas van Fraassen (1970; 1980; 1989) and Elizabeth Lloyd (1994).

Suppes' set-theoretic approach identifies scientific theories with the class of models that are deductively derivable from the fundamental principles of the theory. As van Fraassen put it: 'a model is called a *model of the theory* exactly if the theory is entirely true if considered with respect to this model alone' (1989, p.218). Scientific models thus serve as axiom truth-makers, in that they provide a semantics for the axioms of the background theory that are often expressed as fundamental principles and natural laws. This view is closely related to the concept of models in Tarski's model theory (1953) in which a model of a theory T is a possible realization in which all valid sentences of a theory are satisfied. For Suppes 'the meaning of the concept of model is the same in mathematics and the empirical sciences' and 'the difference [...] is to be found in their use of the concept' (1960, p.289). While mathematicians are primarily concerned with models as possible realizations that satisfy the axioms of theories, when physicists use the term model they are primarily interested with whether the models of the theory correspond well with the relevant experimental data.

In van Frassen's and Lloyd's state-space approach the emphasis shifts on the internal structure of mathematical models in science. In this approach, scientific models are understood as picking out the evolution of physical systems in time by forming trajectories in state spaces. The state space of a physical system can be geometrically interpreted as a Cartesian N-dimensional space whose dimensions are determined by the degrees of freedom of the model. The numerical value for each of the n independent variables of the system at a given point in time comprises an n-tuple, which can then be seen as projecting to a point on the N-dimensional Cartesian state space of the system. The state of the physical system at any given point in time is thus represented by a point in the state space. As the system evolves with time under the influence of various laws determined by the theory, the successive points of each state form the *trajectory* of the system, a curve in the state space representing the system's history of states.

The debate between the Syntactic and the Semantic View on the structure of scientific theories has a long history and has attracted the attention of many philosophers during the last decades (e.g. Suppe, 1977, 1989; Mormann, 2007; Halvorson, 2012; Glymour, 2013; van Fraassen 2014).[3] In more recent years, this tendency of philosophers to reconstruct scientific theories in the formal languages of logic and mathematics has been replaced by a more pragmatic approach that refrains from the strict formalisation of theories and shifts the focus of analysis to further non-formal elements. While acknowledging the importance of the reconstructive axiomatisation and mathematical modelling by the Syntactic and the Semantic View respectively, the Pragmatic View sees scientific theories as internally and externally complex entities, and highlights the importance of various non-formal elements such as the sociology of science and the continuity between theory and experiment. Craver (2002, p.55) nicely portrays this pragmatic approach to theories by saying that

> While these analyses [the Syntactic and the Semantic View] have advanced our understanding of some formal aspects of theories and their uses, they have neglected or obscured those aspects dependent upon nonformal patterns in theories. Progress can be made in understanding scientific theories by attending to their diverse nonformal patterns and by identifying the axes along which such patterns might differ from one another.

The Pragmatic View on the structure of scientific theories was gradually developed by a number of philosophers with similar views, as a response to the Syntactic/Semantic

---

[3] For a comprehensive discussion of the debate see Winther (2021). Interestingly, Sebastian Lutz (2017) argues that this long lasting dispute is essentially about an alleged difference in the dependence of syntactic and semantic approaches on languages of predicate logic that boils down to an illusory distinction between semantic *interpretation* and semantic *representation.*

dichotomy. Winther (2021) identifies the following five core commitments of the Pragmatic View:

1. *Limitations*: The idealised and formalised structure endorsed by the Syntactic and the Semantic View is too weak to ground the predictive and explanatory work that syntacticists and semanticist were aspiring (Cartwright 1983, 1999; Morgan & Morrison 1999; Suárez & Cartwright 2008).[4]
2. *Pluralism*: There is an internal pluralism of theory components (e.g. mathematical concepts, metaphors, analogies, ontological assumptions, values, classifications etc.), as well as an external pluralism of different types of theory (and models) operative in science (Kuhn 1970; Boumans 1999; Hacking 2009).
3. *Non-formal aspects*: The internal pluralism of theories includes several aspects - metaphors, analogies, values, policy views etc. – of a non-mathematical and 'informal' nature that, nonetheless, deserve equal attention (Bailer-Jones 2002; Craver 2002).
4. *Function*: Characterizations of the nature and dynamics of theory structure should pay attention to the users, as well as to their purposes and values (Morrison 2007; Winther 2012).
5. *Practice*: The structure of scientific theories is continuous with practice and experiment, making it difficult to draw a sharp distinction between theory and practice (Hacking 1983; Galison 1987; 1997).

What is another important characteristic of the Pragmatic View is the turn of attention to the importance of scientific models in the analysis of theory structure. As one of the earliest defenders of the Pragmatic View, Cartwright (1999, p.185) calls for a 'reasonable philosophical account of theories' that is 'much more textured and more laborious' than that adopted by the Syntactic and the Semantic Views. In Cartwright's view, theories are abstract mathematical constructs that mainly serve as *tools* for the construction of models (Cartwright et al. 1995; Suarez & Cartwright 2008). As opposed to the Semantic View, in Cartwright's view theories do not contain the good representative models that they produce, rather, they serve as instruments for the construction of models that, in turn, serve as empirical representations of physical phenomena.

---

[4] This feature becomes particularly salient in this thesis in Chapter 4, where the analysis of the LHCb experiment for the construction of data models shows that this practice is extremely complex and hence, less easily formalised than what Suppes' discussion would lead one to believe.

Cartwright's instrumental view of theories is a direct consequence of her Pragmatic View of models developed in her seminal book 'How the laws of physics lie' (1983). In her book, Cartwright argues that the most appropriate units of analysis in order to understand science are models. This is because the natural laws contained in a theory are highly abstract sentences that do not describe reality, but rather, they only apply to the highly idealized objects that are found in models. According to her 'simulacrum account of explanation' (*ibid.*, Ch.8), when fitting theory into data and phenomena, scientists use the existing theoretical frameworks to construct phenomenological models – the simulacra – which, despite the fact that they rely on the background theory, they are not necessarily consistent with it. From these models scientists then derive various phenomenological laws that match the phenomenological behaviour of the physical systems under investigation in greater precision than the abstract theoretical laws of the theory, and thus describe the correct causal stories. To explain a phenomenon is therefore 'to find a model that fits it into the basic framework of the theory and that thus allows us to derive analogues for the messy and complicated phenomenological laws which are true of it' (*ibid.*, p.152).[5]

The independence of models from theories was also emphasised by Morgan and Morrison in their edited collection on 'Models as Mediators' (1999). In their co-authored essay, Morgan and Morrison elaborate their view of scientific models as mediators between theory and reality by arguing that models play a number of essential roles that are separate from the role they play as parts of a theory. Their analysis portrays a picture of scientific models as *autonomous instruments of investigation* embodying elements of independence that can be found in their construction, their function, their ability to represent and their ability to provide knowledge. For Morgan and Morrison, the fact that models include a mixture of elements that partly come from outside the structure of scientific theories as Cartwright showed, provides models with an element of independence from theory during their construction. This independence of models in their construction by virtue of containing elements outside of theory is precisely what enables them to mediate effectively between the theory and the world and serve as a means and a source of knowledge.

This thesis will follow the example of Morgan and Morrison and treat models in physics as *epistemological tools of investigation* whose primary role is to connect theories with experimental data. As Craver noted, the long-lasting study of the structure

---

[5] For a book-length analysis of Nancy Cartwright's views on models and laws of nature see Bovens et al. 2008. For some notable criticisms on Cartwright's views see Forster (1988), Needham (1991), Chalmers (1993), and Earman & Roberts (1999).

of scientific theories and the three main views that stemmed from it have significantly advanced our understanding of the scientific practice, and each of these views has taught us a number of important lessons. Above anything else, the Syntactic View with its persistence on the reconstruction of theories in the language of predicate logic has taught us that the formalisation of scientific practice is not an easy process and will necessarily come with important limitations. Nonetheless, the important dichotomy between theory and observation drawn by the Logical Positivists and the presence of correspondence rules in science is a reminder that any analysis of the scientific practice in terms of theory structure and scientific models must take into account the triangular connection between theory, its users and the natural world. The Semantic view, by identifying theories with their class of deductively derivable models, brought to the surface the importance of models in scientific practice – which was largely ignored by the Syntacticists – and provided, via the state-space approach, a useful framework for understanding how the variables and parameters of mathematical models represent the physical states and properties of physical systems. The set-theoretic approach on the other hand, emphasized the presence of models of theory in science as axiom truth-makers and highlighted – albeit perhaps unintentionally – the important distinction between theoretical models and phenomenological models, which will be discussed in more detail in Sec.1.3. Finally, the Pragmatic View highlighted the importance of various non-formal factors in the study of scientific practice and emphasized the benefits of a more pluralistic and holistic view of science that takes into account a number of sociological and practical problems of a pragmatic nature. The persistence of the Pragmatic view on the study of the experimental life and other practical issues in scientific inquiry can be seen as the predecessor of Philosophy of Science in Practice (Ankeny et al. 2011), a modern approach that aims in answering philosophical questions by closely examining examples of real scientific practice and largely characterises the spirit of this thesis.

In the next section, a novel framework for describing the nature of models in physics will be presented, based on an example from Quantum Electrodynamics and the use of Feynman diagrams therein. This framework characterises models in physics in terms of four core elements: (i) a mathematical framework, (ii) an interpretation, (iii) an ideal system, and (iv) the representational media. It should be noted however, that the presented framework does not amount to a radically new view on the nature of scientific models; in fact, the literature on the nature of scientific models is so vast, that it is unlikely that any radically new accounts on models will be developed in the near future. Rather, the novelty of this framework stems from the fact that it is, to a certain extent, an *integration* of the various existing accounts on models in the literature regarding their mathematical nature, their relationship with physical systems, their ontology and their materialisation. Another important caveat is that this characterisation of models in terms of the above four main features is not a definition of scientific models, nor does it provide the necessary and sufficient conditions for something to count as a model. It has been pointed out numerous times in the literature that models in science can take various forms, and every attempt to define the concept of models in terms of necessary and sufficient conditions is doomed to failure by counterexamples. The aim here is therefore to provide a useful *conceptual basis* for the understanding of models as epistemological tools throughout this thesis. Arguably, these four features can be found to a greater or lesser extent in almost all instances of modelling in physics and an understanding of models in these terms will facilitate the discussions to follow regarding the two main types of models in physics (Sec.1.3) and the role of models in the three major achievements of physics (Sec.1.4).

1.2 Four characteristic features of models

In order to fully understand and appreciate the importance of models in physics, we first need to distinguish them from theories. This is admittedly a difficult and problematic issue, and any attempt to provide a clear-cut distinction will be more or less arbitrary and based on one's understanding of theories and models. For the purposes of our discussion – and throughout this thesis – we shall adopt the rough and ready distinction followed by the physics community that distinguishes theories and models based on their scope and generality. Theories are understood as the primary and comprehensive frameworks that provide the fundamental principles capturing the behaviour of a large class of different yet similar phenomena (e.g. collisions of elementary particles), whereas representational models are understood as specific applications of a theory to a narrow class of phenomena, that might be less certain or incomplete in some aspects.[6]

Theories will be understood here in a Duhemian/Lakatosian manner as consisting of a hard core and a set of auxiliary claims.[7] As Lakatos (1970) famously argued, the core claim of a theory is usually a set of theses that are deemed irrefutable and are often devoid

---

[6] *cf.* Morgan and Morrison (1999, p.12): 'In some cases the distinction between models and theories is relatively straightforward: theories consist of general principles that govern the behaviour of large groups of phenomena; models are usually more circumscribed and very often several models will be required to apply these general principles to a number of different cases.'

[7] This understanding of theories consisting of a core claim together with a set of auxiliary claims is also endorsed by Worrall in his later works (e.g. Worrall 2011; 2014).

of empirical consequences. To derive empirical consequences and predictions from the hard core of a theory one typically needs to include a host of additional hypotheses about the physical system in question, as well as various other principles that are expected to hold across all fields of physics, such as symmetry considerations and conservation laws. For instance, the three laws of Newtonian mechanics that comprise the hard core of the theory, do not by themselves describe the motions of celestial objects. In order to derive the empirical predictions of the theory one needs to employ a host of auxiliary hypotheses and idealisations regarding the conservation of energy, the positions, masses and relative velocities of these objects, as well as techniques of mathematical approximation. Given that the same auxiliary hypotheses can be employed in different physical theories, the hard core of a theory is therefore what differentiates one physical theory from another.

The implementation of auxiliary hypotheses and approximation techniques in order to derive empirical consequences from a theory will be understood throughout this thesis as an *act of modelling*. Often, the primary guiding principle in introducing a model is its mathematical tractability (Redhead 1980, p.147). As we shall see in more detail in Chapter 3, the highly abstract form of physical theories often results in mathematical expressions that cannot be manipulated analytically, and thus, various approximation techniques must be used for the extraction of meaningful results via the construction of mathematically tractable models. Hence, to construct a model of a physical phenomenon is to adjust a background theory to the physical phenomenon by introducing the necessary auxiliary hypotheses and idealisations in order to derive empirically testable results, reflecting the various physical phenomena for which the theory provides knowledge.[8]

Our running example throughout the rest of this chapter will be the construction of a model for the annihilation of an electron and a positron to form a pair of muons ($e^+e^- \to \mu^+\mu^-$) in Quantum Electrodynamics.[9] There are four main reasons for the choice of this particular example. First, this simple reaction, although one of the simplest processes in Quantum Electrodynamics, is one of the most important in high-energy physics, since it is fundamental to the understanding of all reactions in electron colliders and it is often used for the calibration of colliders in large experiments. This example

---

[8] It also helps to understand theories and models as standing in a one-to-many relationship. For instance, the theory of classical mechanics provides the basis for the construction of several models that represent different physical phenomena. These models are constructed by introducing various assumptions adjusted to the specific occasion and phenomena that the models represent. If one removes the specific assumptions of each model, what remains – i.e. the common element in the various models of classical mechanics – is the theory.

[9] The physics for this example and the Feynman method is largely derived from Peskin & Schroeder (1995), a standard textbook for quantum field theory.

therefore represents a non-trivial instance of model construction in one of the most successful theories of modern physics. Second, the particular example nicely illustrates the transition from theory – i.e. quantum electrodynamics – to one of its models. Although it is not easy to identify the exact point at which one departs from the theory and enters the model, the process of constructing a model for the $e^+e^- \to \mu^+\mu^-$ reaction nicely illustrates how one begins from the general fundamental principles of the background theory and ends up with a model of a specific target system that provides concrete and measurable results. Third, this example is closely related to the topic of Chapter 3 in which we shall discuss the wider implications of the perturbative approach in quantum field theory to the relationship between theories and models. The example of the electron positron annihilation to be discussed here comprises a specific case of model construction in quantum electrodynamics and as such, it serves as a good introduction for the more general cases of perturbative quantum field theory to be discussed in Chapter 3. Finally, as we shall see, the choice of this particular example nicely illustrates the four core features by which models will be outlined in this chapter since it is characterised by a solid mathematical framework, its interpretation, an ideal system and its representational media.

Before the construction of the model, theory – i.e. Quantum Electrodynamics – makes its appearance in the form of the QED versions of Maxwell's equations and the Dirac equation, whose solutions (known as Feynman propagators) describe the behaviour of electrons and photons. These equations are derived from the – unique to the QED theory – Lagrangian Density:

$$\mathcal{L}_{QED} = \bar{\psi}(i\gamma^\mu D_\mu - m)\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \tag{1.1}$$

where γ^μ are Dirac matrices, ψ is a bispinor field of spin 1/2 particles and $\bar{\psi}$ is the Dirac adjoint, $D_\mu \equiv \partial_\mu + ieA_\mu + ieB_\mu$ is the gauge covariant derivative with A_μ the covariant four-potential of the electromagnetic field generated by the electron, B_μ the external field imposed by the external source and e the coupling constant equal to the electric charge of the bispinor field, m is the mass of the electron or positron, and $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A\mu$ is the electromagnetic field tensor.

The Lagrangian density of QED is also the starting point for the calculation of time-ordered correlation functions

$$\langle O_1 ... O_n \rangle \equiv \langle 0| \, T[\hat{O}_1 ... \hat{O}_n] \, |0\rangle \tag{1.2}$$

a key mathematical object for the calculation of various observables in quantum field theory, and which, roughly speaking, represents the amplitude of propagation of a particle between n points. Correlation functions in quantum field theory are computed by taking the functional derivatives of the *generating functional* $Z[J]$, which for massless vector fields A_μ and an arbitrary source term J_μ is defined by

$$Z[J_\mu] \equiv \int \mathcal{D}A e^{i\{S[A]+\int J_\mu A^\mu\}} \tag{1.3}$$

where $S[A]$ is the QED Action, which, in turn, is a functional of the generating functional.[10] The generating functional is essentially the quantum analogue of the partition function in statistical mechanics since it tells us everything we could possibly know about a quantum system. As Schwartz (2014, p.262) notes 'the generating functional is the holy grail of any particular field theory: if you have an exact closed-form expression for $Z[J]$ for a particular theory, you have solved it completely'.

The Lagrangian density of QED, along with the set of equations that can be derived by it via the standard procedure through the Euler-Lagrange equation of motion for fields, and the mathematical tools of generating functionals and correlation functions can be understood as comprising the hard core of the theory. In order to derive empirical results from the hard core of the theory one needs to adjust the theory to specific phenomena via the introduction of auxiliary hypotheses and the construction of a model. In Quantum Electrodynamics, as in every quantum field theory, there are two fundamental observables that can be derived from the theory: (i) scattering cross sections σ and (ii) decay rates Γ. Roughly speaking, the cross section of a scattering process expresses the likelihood of obtaining a particular final state, while the decay rate expresses the rate at which an unstable particle decays into a specific final state of two or more particles. The latter is related to the lifetime τ of a particle, which is the reciprocal of the sum of its decay rates into all possible final states. In what follows we shall be concerned with the construction of a model for the calculation of the cross section for the $e^+e^- \to \mu^+\mu^-$ reaction.

Amongst the auxiliary claims of QED there are two 'golden rules' in the forms of formulae for the calculation of cross sections and decay rates in electromagnetic reactions. When it comes to cross sections, in practice, the measurable quantity in experiments concerns not only the final state particles, but also their momenta. This quantity is captured by the differential cross section $d\sigma$ which is a quantity that when integrated over

---

[10] In general, a functional is a function that maps functions to numbers. And just as an ordinary function can be integrated over a set of points x, a functional can be integrated over a set of functions.

any small $d^3p_1...d^3p_n$ gives the cross section for scattering into that region of momentum-space for the final states. In QED the formula for the differential cross section $d\sigma$ for two particles p_A and p_B with corresponding energies E_A and E_B is given by

$$d\sigma = \frac{1}{2E_A 2E_B|v_A - v_B|}\left(\prod_f \frac{d^3p_f}{(2\pi)^3}\frac{1}{2E_f}\right) \times |\mathcal{M}(p_A, p_B \to p_f)|^2(2\pi)^4\delta^{(4)}(p_A + p_B - \sum p_f) \quad (1.4)$$

where the difference $|v_A - v_B|$ is the relative velocity of the beams as viewed from the laboratory frame and E_f is the energy of the final state particles. This is the general formula for the calculation of the differential cross section for every possible process in QED, that is, for every process mediated by the electroweak force. The quantity that differentiates different processes in Eq.(1.4) is the *Feynman amplitude* $\mathcal{M}$. This is a dimensionless quantity that represents the quantum-mechanical amplitude for a particular process to occur and can be thought of as the analogous of the scattering amplitude in non-relativistic quantum mechanics. Different processes have different amplitudes, and thus in order to calculate the differential cross section for a particular type of process such as $e^+e^- \to \mu^+\mu^-$, one needs to adjust the theory to this phenomenon by constructing a model for that process. In what follows, we shall go through the construction of a model for calculating the Feynman amplitude $\mathcal{M}$ and consequently the differential cross section for the $e^+e^- \to \mu^+\mu^-$ reaction, while discussing the four main components of the model.

1.2.1 Mathematical Framework

Models in physics are typically built upon a mathematical framework that is shared with the background theory. The mathematical framework of the model provides the necessary tools for finding exact or approximate solutions to the theoretical equations in order to derive the mathematical consequences of the theory. Hence, the additional structure in the mathematical framework of the model, usually concerns the employment of various approximation techniques in order to make the model mathematically tractable. In the case of QED, the most challenging task is to calculate the Feynman amplitude $\mathcal{M}$ that features in Eq.(1.4) for the various possible processes. The challenge arises from the fact that even for the simplest of QED processes the exact expression of $\mathcal{M}$ is not known, and can only be expressed as a perturbation series in the coupling strength of

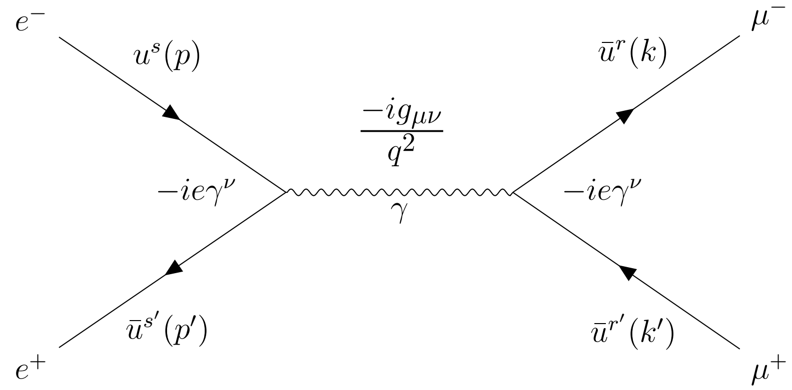

Figure 1.1: Feynman diagram for the first-order term of the Feynman amplitude for the process $e^+e^- \to \mu^+\mu^-$, accompanied by the associated factors derived from the Feynman rules.

the electromagnetic interaction, in which only the first few terms are evaluated.[11] The standard method to achieve this is by using Feynman diagrams. Feynman diagrams provide a delicate way to organize and visualize the perturbation series for $\mathcal{M}$. The idea is that every term of the perturbation series corresponds to a set of diagrams that can be translated directly into a contribution to $\mathcal{M}$ by a set of well-defined rules. In principle, the sum of all terms gives the total value for $\mathcal{M}$, but, in practice only the first few terms are usually calculated, since the complexity of calculations increases dramatically as one proceeds to higher orders and the main contributions to the value of $\mathcal{M}$ come from the first terms anyway. Nevertheless, the evaluation of these expressions is an extremely laborious process that requires the use of powerful computers and it is often the main focus of an entire research project.

Fig.(4.1) illustrates the Feynman diagram corresponding to the first-order term for the amplitude of the $e^+e^- \to \mu^+\mu^-$ reaction, with time flowing from left to right. The diagram depicts the annihilation of an electron-positron pair e^-e^+ into a virtual photon γ which then gives a pair of muons $\mu^-\mu^+$. Feynman diagrams have three components: external lines, internal lines and vertices. Each of these components is associated to an algebraic factor according to the *Feynman rules* of the background quantum field theory. The product of these factors gives the corresponding term in the perturbation series.[12] The associated factors for each component of the specific diagram according to

---

[11] As already mentioned, the implications of this perturbative approach in quantum field theory for the relationship between the background theory and its resulting models will be the main focus of Chapter 3.

[12] For a summary of the Feynamn Rules for QED and other quantum field theories see Peskin & Schroeder (1995, Appendix A.1, pp.801-3).

the Feynman rules for QED can be seen in the figure and when combined they provide the first-order contribution to the amplitude of the process $e^+e^- \to \mu^+\mu^-$:

$$\begin{aligned}\mathcal{M} = \bar{u}^{s'}(p')\,(ie\gamma^\mu)\,u^s(p)\left(\frac{-ig_{\mu\nu}}{q^2}\right)\bar{u}^r(k)\,(-ie\gamma^\nu)\,u^{r'}(k') \\ = \frac{ie^2}{q^2}\left(\bar{u}^{s'}(p')\gamma^\mu u^s(p)\right)\left(\bar{u}^r(k)\gamma_\mu u^{r'}(k')\right) \qquad (1.5)\end{aligned}$$

The actual calculation of the differential cross section for $e^+e^- \to \mu^+\mu^-$ involves a number of further mathematical steps in which the kinematics of the selected frame of reference are applied, the extracted expression for $|\mathcal{M}|^2$ is plugged in the formula for the cross section Eq.(1.4) and an integration over the total phase space is performed. The details of these calculations need not concern us here. Rather, what is of special interest is the *interpretation* one gives to the variables and parameters that appear in the model.

Before we proceed to the interpretation of the mathematical framework, it is important to stress that this framework is the *core* element of all models in physics. As will become evident in the following chapters, the development of the mathematical framework of a model and the presence or absence of mathematical sentences that are in conflict with the mathematical framework of the background theory essentially defines the relationship of the model to the latter. This will become particularly clear in Chapter 3, in which it will be shown that the implementation of non controllable idealizations during the construction of the models in perturbative quantum field theory results in, what we shall call, non-theoretical models.

1.2.2 Interpretation

The interpretation of the mathematical dependencies that appear in a model, is essentially what links the mathematical framework of the model to the physical system it represents. To provide a physical interpretation of the variables and other mathematical objects that appear in the mathematical framework of a model is to attribute a representational relationship between these abstract mathematical entities and the physical quantities they are meant to represent. In more mundane theories, such as classical mechanics, one typically finds a one-to-one correspondence between the variables of the model and the measurable physical quantities they represent. For instance, the mathematical equation for the period of the ideal pendulum – which will be discussed further in Sec. 1.3.1 – expresses a relationship between the period of the pendulum T, the length of the string l and the local gravitational acceleration g. The causal interpretation of this mathematical relationship is that the period of the pendulum is the effect of two causes – the length of the string and the local gravitational acceleration.

The state-space approach of the Semantic View offers a nice conceptual framework for understanding how the variables of a model are related to the temporal evolution of physical systems by forming trajectories in the state-space. This view, however, has its limitations. As one delves deeper into fundamental theories of physics, more and more abstract mathematical objects begin to appear in the mathematical framework of a model, and their physical interpretation, as well as their place in the state-space of models, becomes unclear. Indeed, Duhem (1954, Part II, Ch.I) points out that in theoretical physics one should not expect a clear physical interpretation to every mathematical quantity and operation found in a theory. Rather, as Zahar (1973, p.111) aptly notes, physicists might be led to a new physical insight by trying to find a realistic interpretation of mathematical objects that, at first sight, appear to be devoid of any physical meaning. An example of such an abstract mathematical quantity in QED is the correlation function (or Green's function) mentioned earlier, which is essential for the calculation of the Feynman amplitudes. Correlation functions are vacuum expectation values of time-ordered products of field operators that can be somehow interpreted physically as the amplitude of propagation of a particle – or a field excitation – between two points in spacetime. Nevertheless, this quantity, as many others in quantum field theory, does not correspond directly to a measurable physical quantity; rather, it is only used as a mathematical tool for the development of the mathematical framework that will eventually lead to the empirical predictions of the theory.

In our example, the standard physical interpretation for the variables in Eq.(1.5) is that e is the charge of the incoming electrons, q is the 4-momentum of the virtual photon mediating the interaction, and u, v and $\bar{u}, \bar{v}$ are four-component column-spinors and row-spinors that represent the momentum-space wavefunctions of the initial and final particles with momenta p and k respectively. The mathematical relation expressed by Eq.(1.5) essentially tells us how the (first-order term of the perturbation series of the) amplitude for the process $e^+e^- \to \mu^+\mu^-$ depends on the aforementioned properties of the physical system that our model represents.

In Chapter 4 we shall see how the interpretation of the mathematical framework of the relevant theoretical model is given in terms of the R_K ratio of the branching fractions between two B-decay processes with different flavours of leptons in their final products. This ratio is, in essence, the product of the mathematical interpretation of Lepton Flavour Universality, an abstract theoretical principle of the standard model of particle physics, according to which all leptons (electrons, taus and muons) couple to photons, Z and $W^{\pm}$ in the same way. The interpretation of the mathematical framework of the relevant model ensures that this abstract theoretical principle which is translated into a mathematical ratio within the mathematical framework of the model, corresponds to specific experimental observables related to the yields and the partial decay widths of certain processes of rare B-meson decays.

In the spirit of the Logical Positivists, Cartwright (1983, Ch.7) describes the interpretation of the mathematical framework of a model in terms of – what she calls – the 'model bridge principles' (*ibid.*,p.135). For Cartwright, these principles supplement the traditional bridge principles of the Syntactic view that provide the schemata for getting in and out of the mathematical language of a theory. For instance, while the bridge principles of non-relativistic quantum mechanics tell us that the states of physical systems are represented by vectors and observable quantities are represented by operators, the model bridge principles are the most important ones since those are the principles that tell us how to chose the appropriate Hamiltonian for every concrete and realistic situation that the theory aims to describe.

What is of special interest for the purposes of our discussion, is that, as Cartwright notes, the Hamiltonians that appear in the theory of quantum mechanics are not the Hamiltonians that characterise real atoms and other quantum systems. Rather, they are simply 'model Hamiltonians' that fit the 'highly fictionalized objects' that appear in models. Even in the case of the study of the Hydrogen atom via a quantum mechanical model, says Cartwright, what one studies is not a real Hydrogen atom, since real Hydrogen atoms are never isolated from their environment and their detailed study should necessarily include the effects of the environment in which these atoms are found. What is actually found in the model is a hypothetically isolated two-body system that Cartwright characterizes as a 'mere mental construct' (*ibid.*, p.138). The presence of these mental constructs in models, which we shall call *ideal systems*, is the third main characteristic of models in physics.

1.2.3 Ideal system

Cartwright's (1983) central claim is that the fundamental equations that are found in theories and models do not govern objects in reality. Rather, they govern only highly idealised objects in models, whose simplicity allows their behaviour to be captured by a mathematical framework. A standard characteristic example of such an object is the ideal pendulum in classical mechanics. As the name suggests, the ideal pendulum is a highly idealised version of an actual pendulum in which several idealizations are applied, such as the representation of the bob as a point mass, the elimination of the mass of the string and so on. The exact nature of these ideal objects in models and their representational relationship with actual physical systems has been the main focus of an extensive literature on the ontology of models and scientific representation. In the next chapter, we shall discuss this literature in more detail, and offer a Carnapian solution to the long-standing riddle of the ontology of models. For now, we shall attempt to provide a conceptual framework for understanding these ideal systems, that comes from a rather unexpected place, namely, Meinong's theory of objects.

Amongst the many objects that occupy Meinong's (1904/1960) notorious universe, there is a special type of objects that he calls *incompletely determined objects* (1915/1968, Ch.25). Incompletely determined objects are objects without spatio-temporal existence that are undetermined with respect to at least one property. For instance, the object 'triangle' is perfectly defined as a polygon with three angles and three sides, but nonetheless, it is an incomplete object in that it is neither scalene nor non-scalene, neither isosceles nor non-isosceles and so on. On the contrary, spatio-temporally existing objects such as humans and planets are complete, in the sense that all their properties are determined regardless of whether they are known or not. For Meinong, the main role of incomplete objects is epistemological. They act as mediators between human minds and objects of the external world in order to understand the complicated nature of the latter as complete objects:

> Complete objects are [...] only accessible to us via incomplete objects, which, as we saw, manifest themselves as a sort of aid towards the apprehension of complete objects.[13]

The employment of Meinongian incomplete objects for the description of ideal systems in models is quite illustrative. The complete and messy objects of the physical systems studied by physics have infinitely many determinations which our finite minds and mathematical frameworks can never grasp in their fullness. For this reason, we construct simpler incomplete and finite objects for which our theories can provide precise mathematical descriptions, so that we can understand their behaviour. The ideal pendulum is an incomplete object in that some of its properties, such as the material composition of the bob, its temperature, colour etc., are undetermined since they do not contribute in any way to the investigative purposes of scientists within the theory of

---

[13] Meinong (1915) in Perszyk (1993, p.97).

classical mechanics. By abstracting away all non-relevant properties of a real pendulum one thus arrives at a simple idealized version of the pendulum whose behaviour and properties can be fully and accurately captured by the mathematical framework of the corresponding model.

In our running example in QED, the ideal system that features in the model comprises a simple system in which two point-mass electrons that propagate in space-time annihilate by collision and produce a pair of point-mass muons via the mediation of a 'virtual photon', as illustrated by the Feynman diagram in Fig.(4.1). This is an incompletely determined system, in that not all its quantum numbers are specified. Since the electrons of the model are free electrons, quantum numbers such as the principal quantum number and the azimuthal quantum number related to the states of bound electrons are left undetermined. Moreover, quantities that contribute to the dynamics and kinematics of the reaction but are not known to the modeller are also left unspecified by the model. For instance, there are four possible sets of spin orientations for the electrons and the produced muons, however, the spin orientation of the particles in the model is not specified since it is not known. Instead, the non-zero amplitude for each possible set of orientation is calculated separately and then inserted into the general expression for the cross section of the process. The final result is then obtained by averaging over the four initial-state spin orientations, and summing over the final-state spin orientations.

Let us mention finally, that regardless of how seriously Meinong – and his critics – took the metaphysical implications of his theory of objects, the characterisation of the ideal systems in models as incompletely determined objects presented here does not amount to any sort of ontological commitment about these objects. As will be shown in the next chapter, this is simply a useful 'linguistic framework' that nicely captures the features of the ideal systems in models. What matters is not whether this characterisation is true, but rather, the extent to which it provides, on purely pragmatic grounds, a helpful conceptual framework for understanding the nature of models and their broader role in physics.

1.2.4 Representational media

The fourth and final main characteristic of models is the medium (or media) by which it is expressed and communicated. As we have seen, models in physics comprise various different elements: a mathematical framework that can be expressed via a host of different mathematical expressions, an interpretation of this mathematical framework and an ideal system that is governed by the mathematics of the model. In order to study and communicate these elements, they must somehow be expressed explicitly via the appropriate representational media. Bailer-Jones (2009, pp.2-3) called these media *external representational tools*, noting that when we say that models come in a variety of forms, what we mean is that they employ different representational tools that are not necessarily mutually exclusive. In a similar spirit, Knuuttila (2011) characterises this heterogeneity of representational tools in terms of the 'materiality' of models, a crucial element of their epistemic functioning. For Knuuttila, the representational media by which models are expressed are the material means in which the models are embodied, and it is precisely this material dimension of models that enables their manipulability. Each of these different representational means used by scientists to present a model, affords and limits scientific reasoning in its own characteristic way.

Typical examples of the representational media by which models are expressed in physics are plain language sentences, mathematical equations, plots, diagrams, simulated animations, and even concrete objects. Each of these representational tools is used to convey different bits of information contained in the model, such as information about the target system that the model represents, the various assumptions and idealisations of the model, the relevant laws of nature and other symmetries that constrain the model and so on. In our QED example, we have already employed all of these elements to present the model for the calculation of the cross section for the $e^+e^- \to \mu^+\mu^-$ reaction. We have used plain language to describe the model's relation to the physical system it represents, its connection to the background theory and the interpretation of its mathematical framework. We have also used a mathematical language to express its mathematical framework, and finally, we have used a Feynman diagram to represent a particular mathematical term – the first-order term for the amplitude $\mathcal{M}$ – that features in the mathematical framework of the model.

Feynman diagrams are a prime example of how the representational media of a model can convey different bits of information in a beautiful and succinct manner. The standard convention in all Feynman diagrams is that scalar particles such as pions and the Higgs boson are denoted by dashed lines, vector particles such as photons and the W and Z bosons that mediate interactions are denoted by squiggly lines, and fermions (i.e. quarks and leptons) are denoted by straight solid lines. The arrows on the straight lines denote the direction of charge flow, and not the momentum of the particles. The direction of the momentum – typically denoted by a parallel arrow on a fermion line – is also important however. For fermions and bosons the direction of the momentum is always ingoing for initial-state particles and outgoing for final-state particles, while for internal fermion lines (i.e. propagators), the momentum must be assigned in the direction of particle-number flow, which for electrons it is the direction of negative charge flow.

All this goes to show how – given the standard convention for reading Feynman diagrams – a host of different information can be communicated via a simple diagram such as the one in Fig.(4.1). The value of Feynman diagrams is also found in the fact that besides from being powerful tools and mnemonic devices for making calculations in quantum field theory, they also serve as helpful visualisations for the possible ways in which an interaction with a specific initial and final state can take place in the sub-atomic scale. The standard view in physics is that the lines in the diagram do not represent particle trajectories (as one might see them for instance in a bubble chamber photograph), however, they do provide a useful way of visualising and organising these interactions. Ultimately, what is important is not that the diagrams accurately depict everything that actually happens in the sub-atomic scales of particle physics, but rather, that they allow us to make the right calculations and draw the right conclusions about the empirical consequences of our theories. Indeed, as Feynman diagrams slowly began to take a central role in particle physics, they indicated many unsuspected relations between different physical processes, and suggested intuitive arguments that were later verified by rigorous calculations.[14]

Finally, let us also note that in the literature on the ontology of models which will be the focus of the next chapter, the representational media by which models are expressed and communicated, have occasionally been identified with the ontology of models. That is, in order to answer the question of what models are, some authors have defended the views that (some) models – such as Watson and Crick's metal model of DNA (Schaffner 1969) and Kendrew's plasticine model of myoglobin (Frigg & Nguyen 2016) – *are* physical objects, while others *are* descriptions (Black 2019; Achinstein 1968) and equations.[15] Nevertheless, in the context of the present thesis we shall defend the idea that the physical, linguistic and mathematical objects by which models are materialised are merely the *means* for communicating and manipulating the various aforementioned components of scientific models. As our study in this thesis aims to show, models are multifunctional

---

[14] For more on the representational capacity of Feynman diagrams see Meynell (2008) and references therein. Wüthrich (2010) and Stöltzner (2017) have also argued that Feynman diagrams can be seen as complex mathematical models with representational capacities that go beyond their usefulness as algorithms for performing computations. For a book-length historical approach on the development of Feynman diagrams and their growing role in theoretical physics see Kaiser (2009.)

[15] The view that models *are* sets of equations is a popular view among the scientific community and it is described in Frigg and Hartmann's (2016) review article as an existing view on the ontology of models, however, no references to philosophers explicitly defending this view can be found.

epistemological tools that are much more than a physical object or a set of equations and linguistic sentences, and hence, any informative answer to the question of what models are requires much more than the identification of models with the media by which they are expressed and materialised.

The characterisation of models in physics in terms of their four main characteristics (i) a mathematical framework, (ii) an interpretation, (iii) an ideal system, and (iv) the representational media paves the way towards the main goal this chapter, namely to show how models serve the broader aims of physics. Before that however, it is useful to introduce a very important distinction in physics that will permeate the discussion throughout this thesis, namely the distinction between theoretical and phenomenological models.

1.3 An important distinction

The persistence of the Semantic View to understand theories as 'vending machines' (Cartwright 1999, p.184) that receive empirical input and deliver representational models, brought to the surface the important distinction in physics between theoretical and phenomenological models. Roughly speaking, a theoretical model is a model that can be deductively derived from the fundamental principles of a theory, and hence, the theory is entirely true with respect to the models. A phenomenological model is a model that is primarily constructed with the aim of fitting empirical and experimental data. As a result, a phenomenological model might violate certain laws and principles of the relevant theory with which it is loosely connected, and thus, the theory is not true with respect to the model. It is important, however, to stress that, as Morrison (1999, p.39) notes, the distinction between phenomenological and theoretical models does not mark any particular difference in the intrinsic nature of these models. As we shall see in the following two examples, both models – the ideal pendulum and Bohr's model of the atom – are indeed characterised by the four main features described in the previous section. The difference between theoretical models and phenomenological models is therefore not a difference in their nature, nor is it a measure of the models' accurate representation of their target systems and their ability to function independently of theory. Rather, the differentiating factor is the starting point for the construction of the models; theoretical models are born from the theory, whereas phenomenological models are born from the data. The two examples that follow shall make these remarks clearer.[16]

1.3.1 Theoretical models: The ideal pendulum

A characteristic example of a theoretical model is the ideal pendulum in classical mechanics. The ideal pendulum is an empirically relevant version of a highly abstract theoretical model of classical mechanics, namely the harmonic oscillator. A classical harmonic oscillator can be thought of as an abstract system with a single spatial degree of freedom experiencing a restoring force proportional to its displacement from an equilibrium position. To make this abstract model relevant to real oscillating systems such as pendulums, a number of additional features must be introduced.

In its simplest version, the empirically relevant model of the ideal pendulum is described as a point-particle of mass m in a homogenous gravitational field g, connected to a fixed point by an inextensible and massless string of constant length l, which is free to move in a circle of radius l in a vertical plane (Fig.1.2). These assumptions imply that the only force acting on the mass is gravity, establishing this way that the system behaves as a simple harmonic oscillator subject to a continuous restoring force $F = -mg$. The model is specified by a family of six variables, each corresponding to a measurable physical quantity, i.e. an observable: (1) the angle θ between the string and the vertical (which represents the system's single spatial degree of freedom), (2) the length l of the string, (3) the mass m of the bob, (4) time t, (5) the period T of the pendulum, and (6) the gravitational acceleration g.[17]

Once all basic assumptions are in place, applying Newton's second law within the mathematical framework of the model provides the pendulum's equation of motion

$$\frac{\mathrm{d}^2\theta}{\mathrm{d}t^2} + \frac{g}{l}\sin\theta = 0 \tag{1.6}$$

which specifies a response function according to which a measurement of θ will yield a specific value θ_1, given a state s_1.[18] However, Eq. (1.6) is a second order non-linear

---

[16] As a precaution, let us note here that in Chapter 4 we shall use – in lack of a better term – the broader concept of 'models of theory' to denote a wider class of models including both theoretical and phenomenological models, as well as the non-theoretical models to be discussed in Chapter 3. As will become clearer in that chapter, the reason behind the choice of this perhaps confusing terminology is to distinguish between the models coming from the theoretical side of physics and the data models coming from the experimental results.

[17] The observables of a model may represent either *base quantities*, i.e. conventionally chosen physical quantities that cannot be expressed in terms of other quantities (θ, l, m, and t), or *derived quantities*, i.e. physical quantities that are defined in terms of (at least some of) the base quantities (T and g).

[18] In the language of the state-space approach, the set S of possible states specified by the model of

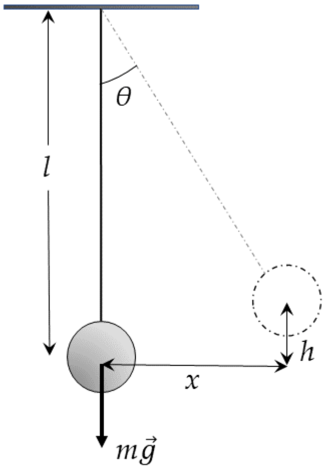

Figure 1.2: The ideal pendulum.

differential equation for which no exact analytic solution can be found, and therefore, to make the model mathematically tractable various mathematical idealizations must be applied.

The standard way to achieve this is by introducing the assumption that the pendulum only swings in small angles for which we know that

$$\sin\theta \approx \theta \tag{1.7}$$

The small angle approximation yields a new linear equation of motion

$$\frac{\mathrm{d}^2\theta}{\mathrm{d}t^2} + \frac{g}{l}\theta = 0 \tag{1.8}$$

which has an exact solution of the form

$$\theta = \theta_0 sin(\omega t + \phi_0) \tag{1.9}$$

where θ_0 is the angular amplitude of the swing, ω is the angular frequency, and ϕ_0 is the initial phase angle depending on the initial conditions. The linearised approximation also gives the mathematical relation

$$T = 2\pi\sqrt{\frac{l}{g}} \tag{1.10}$$

---

the pendulum has as its members all possible combinations of the numerical values for each one of the six observables of the model, with respect to their allowed range specified by the laws of the theory. Each state s_i therefore stands for a possible configuration of the system, and is represented in the state space by a point with coordinates $[\theta_i, l_i, m_i, t_i, T_i, g_i]$.

for the period of the pendulum which, in accordance with Galileo's famous conclusion of isochronism, is constant.

A further assumption that ensures the mathematical tractability of the model is the assumption of a homogenous gravitational field. Newton's law of universal gravitation $F = GMm/r^2$ combined with his second law $F = mg$ gives

$$g = \frac{GM}{r^2} \tag{1.11}$$

where G is the universal gravitational constant, M is the mass of the Earth and r is the distance from the Earth's centre to the point for which the calculation is carried out – in this case is the centre of mass of the bob. The value of the gravitational field is therefore, according to the background theory, inversely proportional to the square of the distance between the centres of mass of the two bodies.

The justification for the assumption of a homogenous gravitational field is achieved by showing that

$$\frac{GM}{r_{min}^2} - \frac{GM}{r^2} \approx 0 \tag{1.12}$$

where the minuend represents the value of the gravitational field at the lowest point, for which the distance between the centre of mass of the bob and the centre of the earth is r_{min}. The subtrahend represents the value of the gravitational field at the various points along the trajectory of the bob as the pendulum swings, for which the square of the distance between the centres of mass of the two bodies is given by $r^2 = (r_{min} + h)^2 + x^2$, where x and h represent the displacement of the bob in the horizontal and vertical axes respectively (see Fig.(1.2)). Given that for small oscillations $x \ll r_{min}$ and $h \ll r_{min}$, the difference in (1.12) approximates zero and thus the homogeneity of the gravitational field becomes a good approximation to the experimental setup.

The astute reader may notice here that, strictly speaking, the linearised equation of motion of the ideal pendulum (Eq.(1.8)) is not deductively derivable from the mathematical framework of the theory of classical mechanics, since it only follows from the somewhat arbitrary introduction of the small angle approximation. In our discussion so far, we have defined theoretical models as models that are deductively derivable from theories and for which the theories are true. However, while this is true for the highly abstract model of the harmonic oscillator, it does not seem to be the case for the linearised version of the ideal pendulum. This should not come as a surprise. Purely theoretical models such as the harmonic oscillator are usually *physically irrelevant mathematical artefacts* with no tangible connection to the physical systems and phenomena they are supposed to represent. For this reason, the introduction of empirically relevant models is often needed in order to compare the predictions of the theory with the outcomes of experimental measurements.

The construction of empirically relevant (theoretical) models, such as the linearised version of the ideal pendulum, necessarily comes with the introduction of various assumptions and idealisations. The exact nature of these idealisations and their consequences on the relationship between theory and models will be discussed in detail in Chapter 3. For now, is suffices to say that what is important is that the introduced idealisations in the model of the ideal pendulum, are, what Sklar (2000, pp.44-5) calls, *controllable*. In other words, what makes a model a theoretical model is that the construction of the model begins from the theory and the various introduced assumptions and idealisations can be physically or mathematically justified. Insofar as we have good mathematical and physical reasons to justify Eq.(1.7) and Eq.(1.12), the linearised ideal pendulum is still a theoretical model, although, strictly speaking, the theory is not entirely true of the model.

The upshot is that the starting point for the construction of a theoretical model comes from the fundamental principles of the background theory – in this case Newton's second law – and that the model is constructed in order to test the empirical consequences of the theory in real pendulums. We can also see that the model is indeed characterised by the four main features of models. The mathematical framework of the model is derived from the mathematical framework of classical mechanics, but it is also enriched with the two aforementioned mathematical approximations (Eq.(1.7) and Eq.(1.12)) in order to make the model mathematically tractable. The mathematical framework of the model is then interpreted in a particular way – i.e. each variable is linked to a measurable physical quantity – which gives rise to an ideal incomplete system. The model is then expressed and communicated by various representational means, such as physical language, mathematical language and various diagrams such as Fig.(1.2).

1.3.2 Phenomenological models: Bohr's model of the Hydrogen atom

The two defining characteristic features of phenomenological models is that their construction begins from the experimental data and that they often incorporate claims that are in conflict with the relevant background theory. A nice historical example of a phenomenological model in physics comes from Bohr's model of the hydrogen atom. Bohr devised the planetary model of the hydrogen atom in 1913 in order to explain the stability of matter and the discrete nature of the spectral lines of hydrogen, two physical phenomena that the earlier proposed theoretical model by Rutherford failed to explain. Rutherford's model was built based on the observations from his famous experiments with gold foils, while maintaining a perfect consistency with the basic principles of the then available theory of classical electrodynamics. In Rutherford's version, the electrons of the atom orbit the positively charged nucleus under the influence of a Coulomb electrostatic attraction, similarly to the way the planets orbit the sun attracted by gravity. However, if the electrons were moving in a circular orbit around the nucleus, they should exhibit radial acceleration and hence, according to the basic principles of classical electrodynamics, they should continuously emit radiation leading to a constant loss of energy. This would then force the electrons to a spiral trajectory onto the nucleus, eventually causing every single hydrogen atom to collapse. Moreover, the continuous radiation of the electrons due to their acceleration was at odds with the discrete nature of the emission spectrum of the hydrogen in visible light, which was already observed in earlier experiments and was mathematically expressed by the Swedish mathematician Johann Balmer.

In order to explain these phenomena, Bohr modified the existing theoretical model by introducing the following two postulates: (i) the only possible orbits are those for which the angular momentum of the electrons is an integer multiple of the reduced Planck constant $\hbar$ and (ii) radiation is emitted or absorbed only when an electron 'jumps' from one orbit to another. The introduction of these two postulates allowed for an explanation of the stability of matter since electrons could now be found in a 'ground state' where, despite their acceleration, no radiation was emitted and thus, there was no loss of energy. As is well known, Bohr's model was also found to be in remarkable agreement with the experimentally derived values of the radius of the atom and the energy of the ground state, and gave a solid explanation both of Balmer's formula for the spectral lines and the value of the Rydberg constant therein.[19]

Bohr's model of the hydrogen atom is a typical example of a phenomenological model whose construction is primarily guided by empirical observations, with the aim of matching the predictions of the model with the available experimental data. This technique is quite common in physics especially in those cases where the existing theoretical models are either too difficult to be solved analytically, or they simply fail to accommodate the corresponding experimental outcomes. What makes Bohr's model a non-theoretical

---

[19] For further details on the derivation of the Balmer series and the Rydberg constant from Bohr's model see Bokulich (2011, Sec. 4).

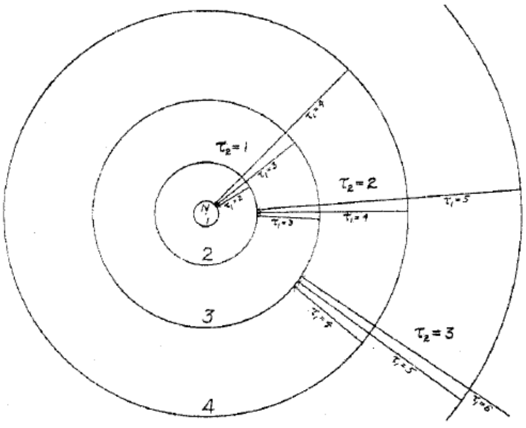

Figure 1.3: Early depiction of Bohr's model of the Hydrogen atom by Harkins and Wilson (1915). The diagram shows the nucleus N of the atom in the centre, surrounded by the allowed orbits 1,2,3,4 of the electron in the various allowed states of motion.

model is the fact that some of its core assumptions are in direct conflict with the basic principles of the theory within which the model was introduced. For instance, Bohr's second postulate – that the electrons do not radiate while in orbit – contravenes the fundamental principle of classical electrodynamics which says that any charged accelerated particle must emit radiation according to Maxwell's equations.[20] This is not to say that Bohr's model is completely independent of the theory however. Like all models in physics, Bohr's model of the hydrogen atom was indeed partially inspired by the background theory of classical electromagnetism: the electrons are charged particles subject to a Coulomb potential from the nucleus, they are assigned an angular momentum, and they follow a continuous periodic trajectory. Nonetheless, it is a non-theoretical model, in the sense that the introduced idealisations via the two postulates are not controllable, given that at the time the model was introduced, there was no theoretical justification for the fact that some of the basic principles of the supporting theory were infringed.

Just like the ideal pendulum – and most models in physics – Bohr's model of the hydrogen atom is also characterised by the four main features introduced in Section 1.2, indicating that the difference between theoretical and phenomenological models is not a difference in their intrinsic nature. The ideal system of the model comprises of two particles, the heavy nucleus and the orbiting electron, between which there is an attractive force corresponding to the potential energy function

---

[20] For a rigorous discussion of the inconsistencies of Bohr's model with classical electrodynamics see Vickers (2013, Chapter 3).

$$V(r) = -\frac{Ze^2}{r} \tag{1.13}$$

where r is the distance between two particles, and Ze represents the nucleus' electric charge. The mathematical framework of the model is expressed by a set of equations including the potential energy function and the equations of motion of the system. However, the crucial difference with theoretical models is that the mathematical framework of Bohr's model is not derived solely from the mathematical framework of the theory with the addition of approximations. Rather, Bohr's two postulates induce further mathematical constraints on the orbits

$$r_n = \frac{\hbar^2}{me^2} n^2 \tag{1.14}$$

velocities

$$v_n = \frac{e^2}{\hbar} \frac{1}{n} \tag{1.15}$$

and allowed energy levels of the orbiting electrons

$$E_n = -\frac{me^4}{2\hbar} \frac{1}{n^2} \tag{1.16}$$

where $\hbar$ is the reduced Planck constant, m is the mass of the electron, e is the electron charge and $n = 1, 2, 3, ...$ represents the energy level of the orbiting electrons. The interpretation of the mathematical framework of the model generates an ideal system comprising of a two-body system in which the electron orbits the nucleus of the hydrogen atom in predefined quantised orbits, and it can of course be represented by various representational media, such as the above equations and the familiar diagrams of the planetary model for the hydrogen atom. For instance, Fig.(1.3) shows a depiction of the model by Harkins and Wilson dating back to 1915, which, according to Kragh (2012, p.61), is probably the first diagrammatic depiction of Bohr's model of the Hydrogen atom.

1.3.3 A third alternative?

The distinction between theoretical and phenomenological models in physics is fundamental to the understanding of the scientific method and the evolution of theories. On the one hand, one derives models from the available theories to validate these theories, explore their empirical consequences and predictions, and to explain phenomena within the framework of the theories. On the other hand, one constructs models based on experimental data that cannot be accommodated by existing theories. And just like Bohr's model, when successful, these phenomenological models can be the starting point for the refinement of existing theories or even a paradigm shift. It should be noted however, that rather than highlighting the existence of two different types of models, the distinction between theoretical and phenomenological models is better understood as denoting the two ends of a continuous spectrum. On the theoretical end, one finds highly abstract and purely theoretical models such as the simple harmonic oscillator, and on the phenomenological end, one finds models that are completely detached from any physical theory. The historical development of the model for the hydrogen atom shows that scientists may begin with a theoretical model and then introduce the necessary modifications in order to account for certain phenomena that the theory fails to explain. This process of adding empirical elements and ad-hoc hypotheses to a model of a theory, detaches it from the theoretical end of the spectrum and shifts it to the phenomenological side. When theory is refined in order to accommodate the phenomenological model, the model then again shifts back to the theoretical side of the spectrum and becomes a theoretical model of the new and refined version of the theory.

In a relatively recent paper, Reutlinger et al. (2020) discuss the nature of *autonomous toy models*, implying the existence of a third type of models that are neither theoretical nor phenomenological. Autonomous toy models are models that are not embedded in theoretical frameworks, but are not constructed on the basis of existing experimental data either. Their main example is Schelling's (1971) model of racial segregation, an agent-based model in sociology that illustrates how racial segregation may occur in a city where the inhabitants have a slight preference for neighbours of the same race. The idea is that the model is not derived from any well-established theoretical framework, nor is constructed on the light of prior existing data, and thus, it exhibits a form of autonomy from the background theory. To support their claims, Reutlinger et al. provide two further examples – the Lotka-Volterra model in biology and the DY-model in econophysics. It is, however, questionable whether such autonomous models that are completely detached from a theoretical framework can be found in physics. When the discussion comes to physics, Reutlinger et al. discuss the MIT bag model for quark confinement noting that although it is is not a model of the background theory of quantum chromodynamics, it is, nonetheless, inspired by and non-trivially associated with the theory via a connecting story, as argued by Hartmann (1999).

In an older work, Redhead (1980) also discusses the possibility of a third alternative type that he calls *floating models*. A floating model is 'a model which is disconnected from a fundamental theory T by a computational gap in the sense that we cannot justify mathematically the validity of the approximation being made but which also fails to match experiment with its own (model) predictions' (Redhead 1980, p.158). As an example, Redhead briefly discusses Elliott's (1958) SU(3) model of the nucleus, in which the initially imposed SU(3) symmetry of the wavefunctions is strongly broken in order to achieve a match between theoretical and experimental energy levels. Although the resulting model is not a model of the theory, the value of floating models according to Redhead is that they can be thought of as capturing essential features of the underlying theory, leading to new insights about the possible ways of refining existing theories.

In Chapter 3, we shall discuss the importance of non-phenomenological models that are also detached from the theory, within the framework of perturbative quantum field theory. In particular, we shall examine the process of constructing a model for scattering processes in perturbative quantum field theory via the methods of regularization and renormalization. The careful analysis of this process will show that the resulting models are indeed detached from the background theory by a *computational gap* (to use Redhead's terminology) in that some of the mathematical steps used in the development of the mathematical framework of the models concerning the treatment of infinities in the perturbation series, cannot be justified mathematically or theoretically. The main difference between these models and Redhead's floating models however, is that, despite being detached from the theory by a computational gap, models of perturbative quantum field theory exhibit a remarkable agreement with experimental data, which raises some interesting questions about the relationship between theories and models and the empirical adequacy of the background theory. We shall postpone the discussion of these interesting questions until Chapter 3. For now, let us examine the role of models in achieving the main aims of physics and see how our characterisation of models by the four above mentioned features and the distinction between phenomenological and theoretical models fit within the discussion. The next and final section of this chapter comprises three parts in which the three different major achievements of physics – economy, explanation, and prediction – are discussed. Each part begins with a brief review of the existing literature on these topics and leads to a discussion of the role of models therein.

1.4 How models serve the aims of physics

The question ‘What is the aim of physics?’ is not a simple one. Arguably, this question lies at the heart of the time-honoured debate on scientific realism. The realist-inclined philosopher, and perhaps the theoretical physicist, insists that physics aims at truth; that is, it aims in grasping the fundamental reality of matter as accurately as possible by offering true descriptions of the real nature of matter and genuine explanations of why certain phenomena occur the way they do at the most fundamental level. The antirealist-inclined philosopher, and perhaps the experimentally oriented physicist, might assert that the aim of physics – or to be more precise, the aim of some physicists – might indeed be the search for truth, but argue that this task is in fact unachievable. Hence, the antirealist insists that the aim of physics does not necessarily relate to the search for truth. Rather, our physical theories have the more modest aim of ‘saving the phenomena’ by offering possible but not necessarily genuine explanations of data patterns, as well as results and predictions that are in agreement with the experimental data.

To the former, Pierre Duhem responds that the genuine explanation of phenomena cannot be the aim of physics. To explain, says Duhem, ‘is to strip reality of the appearances covering it like a veil, in order to see the bare reality itself’ (1954, p.7). The removal of this veil requires to proceed deeper and ‘underneath those appearances’ in order to reveal the true nature of matter and show that our perceptions – or our experimental data in more modern terms – are produced *as if* the reality is what the theory asserts. However, for such an inquiry to be possible one first needs to provide an answer to (i) whether there indeed exists a material reality under the sensible appearances that are revealed in our perception and (ii) what the true nature of the elements that constitute such a reality is. The resolution of these questions however, transcends the methodology of physics, according to Duhem, and therefore, if the aim of physical theories is to explain the experimental results, theoretical physics loses its status as an autonomous science, and becomes subordinate to metaphysics. Duhem then concludes that ‘our theories have as their sole aim the economical condensation and classification of experimental laws; [and hence] they are autonomous and independent of any metaphysical system’ (*ibid.*,p.219).

To the latter, Henry Margenau objects that ‘the driving force in physical investigation can not be exclusively the fun or the profit of prediction’ (1935, p.54). For there are indeed numerous important physical theories which have failed to produce any noteworthy predictions, such as the theories of molecular binding and of ferromagnetism which, at the time of Margenau's paper, where 'so far ahead of "explanation" that a long time is likely to elapse before any significant purely theoretical predictions of new phenomena will be made' (*ibid.*). In modern days, one might argue that string theory is a characteristic example of a physics theory that has not yet produced an experimentally testable prediction but it, nonetheless, occupies a significant part in the enterprise of physics. Margenau's verdict is therefore that 'the physicist's business involves something more than prophecy, something that makes his experience peculiarly coherent and produces an internal fitness' (*ibid.*).

These sketches of Duhem's and Margeneu's arguments are two simple examples from a long list of well-thought and sophisticated arguments for and against the various dimensions of the view that physics aims at truth, as a result of the long-standing debate on scientific realism. To avoid any objections of this kind we shall pursue a more modest task. Instead of arguing that physics has a specific aim (or aims), we shall follow the opposite direction and ask 'What do we actually accomplish by doing physics?'. That is, following a more pragmatic approach, we shall proceed in identifying what we consider to be the most important practical *outcomes* or *achievements* of physics, and then identify the role of models in achieving these goals. Given the vast literature on the nature of models, their representational capacities and their place in the debate on realism, it is surprising that models have not been explicitly linked with the broader aims of physics (the topic of scientific explanation is perhaps the only exception). Rather, the usual approach is to study the various specific purposes for which models are built – e.g. to test a theory, to explain a specific physical phenomenon, for educational purposes, to accommodate data that cannot be explained by the theory etc. – and evaluate their role therein.

In what follows, it will be argued that the three most important outcomes of physics are (i) the economic description of nature, (ii) explanation, and (iii) prediction. In short, *economic description* refers to the fact that physics provides the tools for organizing and classifying physical phenomena and physical entities in a helpful manner, *explanation* refers to the fact that physical theories provide genuine and/or possible explanations for why certain phenomena occur the way they do and why certain patterns of data are produced in experiments, and *prediction* refers to the fact that, through physics, scientists derive results concerning 'novel phenomena' as well as previously known phenomena that could not be accommodated by existing theories. Notwithstanding deeper metaphysical questions regarding the truthfulness of scientific theories, the idea here is that these three achievements of physics are common ground between realists and antirealists, in that neither side of the debate denies that physics does indeed succeed in producing these outcomes. Our ultimate goal is to show how and to what extent the art of modelling contributes to the fulfilment of each one of these three achievements using the presented four-features framework.

1.4.1 Economic description

The idea that physical theories provide economic descriptions of nature goes back to Ernst Mach's principle of the *economy of thought* expressed in his 1882 lecture on 'The economical nature of physical enquiry' (Mach 2014) and later incorporated in his 1883 book on 'The science of mechanics' (Mach 2013). According to this doctrine, the scientific laws and other abstract mathematical terms in our physical theories are tools for representing, organising and compiling our experiences in our thought by means of the most economical possible ways, in order to be able to efficiently control and reproduce physical events. Heavily influenced by Hume, Mach endorsed a conception of nature according to which the natural world does not consist of stable objects and laws but rather of a mosaic of individual and constantly changing sensations which he called *elements*. Some of these elements were human and animal sensations like colour, sound and tastes, while others were physical magnitudes such as pressures and point potentials. Mach's world consists of a constant and bewildering stream of events which never recur in their full capacity and for this reason, human minds are incapable of grasping it in every detail. Science acts as a tool for organising and accumulating the contents of our thought making this way knowledge about the external world attainable.

In physics, this economy of thought is achieved, according to Mach, by the implementation of natural laws and the mathematisation of theories. In Mach's Humean conception of the world, laws of nature do not imply necessary connections between different existences, rather, they are efficient tools for describing certain patterns of events that occur in the world.[21] They are single expressions providing instructions for the mental reconstruction of a great number of individual facts which would otherwise require a tremendous amount of time and energy to be reproduced:

> ...instead of noting individual cases of light-refraction, we can mentally reconstruct all present and future cases, if we know that the incident ray, the refracted ray and the perpendicular lie in the same plane and that

---

[21] See Bhogal (2020) for a recent comprehensive discussion of the Humean conception of natural laws.

> $sin\alpha/sin\beta = n$. Here, instead of the numberless cases of refraction in different combinations of matter and under all different angles of incidence, we have simply to note the rule above stated and the values of n – which is much easier. The economical purpose is here unmistakable. In nature there is no *law* of refraction, only different cases of refraction. The law of refraction is a concise compendious rule, devised by us for the mental reconstruction of a fact... (Mach 2013, pp.485-6, emphasis in the original.)

The laws we find in physical theories such as Newtonian mechanics and thermodynamics can therefore be seen as simple propositions that allow us to reconstruct a great number of individual events. They are not, however, the only tool via which the economy of thought is accomplished in physics. Mach continues by saying that 'mathematics is the method of replacing in the most comprehensive and *economical* manner possible, *new* numerical operations by old ones done already with known results' (*ibid.*, p.487, emphasis in the original) and hence, 'the sciences most highly developed economically are those whose facts are [like mathematics] reducible to a few numerable elements of like nature'(*ibid.*, p.486).[22]

As a highly mathematical science, physics therefore achieves a high degree of economy with the mathematisation of relations between the physical elements of its theories:

> Physics shares with mathematics the advantages of succinct description and of brief, compendious definition, which precludes confusion, even in ideas where, with no apparent burdening of the brain, hosts of others are contained. Of these ideas the rich contents can be produced at any moment and displayed in their full perceptual light. (Mach 2014, p.197)

Using the example of falling bodies, Mach shows how the mathematical treatment of a falling body saves us from the labour of reconstructing the full range of all possible

---

[22] It should be noted however, that Mach's principle of the economy of thought is not equal to the principle of quantitative parsimony discussed, for instance, in Nolan (1997), Baker (2003), and Jansson & Tallant (2020). As stated by Nolan, quantitative parsimony is the view that, following Occam's razor, our scientific theories should aim to minimize the number of kinds of entities they postulate. It is a view about the ontology of scientific theories, and is regarded by its proponents as a theoretical virtue in that it makes our scientific theories more explanatory and brings them closer to truth. While Mach was sympathetic to the idea of accommodating empirical facts in the most economically possible way – and thus one may argue that the most economic way is the one that postulates the fewest entities – he had little to say about the value of ontological parsimony as a guide to truth (*cf.* Banks 2004, p.25). Mach's principle of the economy of thought is not a guide for the ontology of scientific theories. It is pragmatic virtue of science; it is what makes knowledge about the infinitely complicated mosaic of elements that constitute the natural world possible.

motions of falling bodies one by one. The mathematical formula serves as a 'complete substitute for a full table of motions of descent, because by means of the formula the data of such table can be easily constructed at a moment's notice without the least burdening of the memory' (*ibid.*, p.193). In modern terms, perhaps the most characteristic example of how the mathematical framework of a theory serves as an organizing tool for the economy of thought comes from the famous Lagrangian of the standard model which contains in a single equation all the kinetic terms, mass terms, coupling terms, and the Higgs mechanism for all the fundamental particles. The Lagrangian of the standard model serves as the starting point for a host of useful calculations in quantum field theory and applied high-energy physics regarding the equations of motions for particle collisions, path integral calculations, and calculations of experimentally useful quantities such as cross sections. The mathematical formulation of the theory of the Standard model in terms of a single Lagrangian can therefore be thought of as a compact description of all the theoretical and experimental advances associated with particle physics since the middle of the 20th century, which would otherwise require thousands of pages to be expressed in full detail.

It is not hard to see how the same idea can be applied to models. While Mach's discussion is mostly focused on the role of laws within theories, his remarks can be easily applied to models as well. Just as the mathematical framework of the theory for a falling body saves us from the labour of reconstructing all possible motions of a falling body, the interpretation of the mathematics which gives rise to the incomplete ideal systems in models – discussed in Secs. (1.2.2) and (1.2.3) – allows us to construct a single model for a host of different physical falling systems, saving us from the labour of constructing a different model for each different type of falling body in nature. Recall that by giving an appropriate interpretation to the mathematical framework of the model and by abstracting away all non-relevant properties of a physical system one constructs an incompletely determined system which is characterised only by a handful of useful properties. Given that these properties can be found in a variety of actual physical systems, the ideal object is 'embedded' in these systems and can thus be used to probe their behaviour. Indeed, the highly abstract model of the classical harmonic oscillator serves as the starting point for the study of a variety of oscillating systems in nature, ranging from simple pendulums to chemical bonds and helium-neon lasers.

A strong advocate of Mach's views on theories as tools for the economy of thought was Pierre Duhem.[23] In his classic book 'The aim and structure of physical theory'

---

[23] Arguably, another proponent of Mach's views on the aims of science was Albert Einstein: 'The

(1954), Duhem devotes a large part of the second chapter arguing that the economy achieved by reducing the innumerable concrete facts of nature into laws discussed by Mach, is increased even more when these laws are further reduced into the fundamental principles of a theory:

> ...instead of a great number of laws offering themselves as independent of one another, each having to be learnt and remembered on its own account, physical theory substitutes a very small number of propositions, viz., fundamental hypotheses. The hypotheses once known, mathematical deduction permits us with complete confidence to call to mind all the physical laws without omission or repetition. (Duhem 1954, p.21)

Duhem's idea is that the laws of a theory can be reduced into a small number of fundamental principles, since these principles can then be used to derive all the laws of the theory by means of mathematical deduction. A characteristic example of how fundamental assumptions in a physical theory are used to derive its laws comes from the two postulates of special relativity as stated by Einstein in the introduction of his famous 1905 paper 'On the electrodynamics of moving bodies' (Einstein 1905). Einstein's two fundamental assumptions, namely that (i) the laws of physics are valid for all frames of reference and (ii) that light is always propagated in empty space with a constant velocity which is independent of the state of motion of the emitting body, form the basis on which Einstein derived Lorentz invariance, which in turn serves as the basis for the derivation of a number of consequences of the theory such as time dilation and the relativity of simultaneity. Hence, what laws are for facts in Mach's view, theories are for laws for Duhem.

In discussing Mach, Duhem takes one step further and adds that physical theories also serve as *classifications* of natural laws, and perhaps, of nature. Theory, says Duhem, by developing a ramified picture of the deductive reasoning from fundamental principles to natural laws, establishes at the same time an order and classification of these laws, which not only makes the laws easier to handle and more useful, but also brings with it a sense of beauty. To use Duhem's clever analogy, 'theory gives, so to speak, the table of contents and the chapter headings under which the science to be studied will be methodically divided, and it indicates the laws which are to be arranged under each of these chapters' (1954, p.23).

---

grand aim of all science is to cover the greatest possible number of empirical facts by logical deductions from the smallest possible number of hypotheses or axioms' (Einstein et al. 1988, p.282).

More importantly however, this handiness and beauty of the ordering of laws in a theory also brings to the scientist, according to Duhem, a strong belief that the theory is grasping a natural *classification*. For Duhem however, these classifications do not necessarily reflect real classifications in nature. Rather, in the spirit of Mach's pragmatism, they are 'purely ideal connections' in the minds of the naturalist serving as a useful synoptic table which summarizes her conception of physical facts. A natural classification is for Duhem, 'a group of intellectual operations *not referring to concrete individuals* but to *abstractions*, species; these species are arranged in groups, the more particular under the more general' (*ibid.*, p.25, emphasis added). What cultivates the strong belief that the classification is natural, i.e. that it corresponds to a categorisation that reflects the structure of the natural world rather than the interests of scientists, is an *act of faith* stemming from the admirable ability of the theory to bring together a host of innumerable and complicated facts and condense them in a neat and ordered fashion:

> ...when we see in the plan drawn by these hypotheses a vast domain of optics, hitherto encumbered by so many details in so confused a way, become ordered and organized, it is impossible for us to believe that this order and this organization are not the reflected image of a real order and organization. (*ibid.*, p.26)

Duhem's observations on theories as classifications are reflected in contemporary philosophy by the literature on natural kinds, especially in the philosophy of biology. A large part of the current debate on natural kinds concerns the question whether the various natural kinds postulated by theories correspond to a grouping that reflects the structure of the world as the realists assert (e.g. Boyd 1991) or to a grouping that reflects the interests and actions of scientists as per the conventionalist view (e.g. Hacking 2007).[24] It is not the purpose of this chapter to examine this question however. Rather, what is of interest for our purposes is that one of the major achievements of physics is that it indeed provides an elegant description of nature by identifying the common features of various physical systems, regardless of whether this classification corresponds to a real classification of nature or not. The mathematical framework of the standard model, for instance, provides the means for the classification of all elementary particles in two broad categories (fermions and bosons) and their further division into sub-categories according to their flavour and other properties. Bosons are divided in vector bosons (the

---

[24] For a comprehensive review of the literature on natural kinds see Bird & Tobin (2008) and references therein.

W and Z bosons, also known as the force carriers) and scalar bosons (the Higgs boson), while fermions are divided to leptons (electrons, muons and taus) and quark pairs (up-down, charm-strange, top-bottom) and so on.

This useful classification is not only achieved by theory however. By isolating the common features of different physical systems, scientific models provide a helpful classification of physical systems in terms of similar patterns in their behaviour. This idea is profoundly evident in Batterman and Rice's (2014) work on minimal models. Minimal models are caricature models that aim to capture in the most economical way possible the essential physics of a host of various physical systems that belong in the same *universality class*. According to Batterman and Rice, physical phenomena belong in the same universality class insofar as they exhibit the same patterns of behaviour, e.g. by exhibiting a continuous behaviour as fluids.[25] Batternman and Rice's argument is that the explanatory capacity of minimal models stems from the fact that the models belong to the same universality class as the physical systems. Leaving aside the details of their argument, what is important for our discussion is that these minimal models serve as classifications for the various physical systems they represent. The harmonic oscillator demarcates the universality class for all kinds of oscillating physical systems, and the QED model for the process $e^+e^- \rightarrow \mu^+\mu^-$ demarcates the universality class for all possible reactions including a pair of electrons in their initial state and a pair of muons in their final state. Different universality classes therefore classify different types of physical systems that can be grouped together as the target systems of a minimal model.

Another way to appreciate the role of minimal models in the classification of nature is to understand them in terms of the presented four-features framework. Insofar as the minimum set of properties attributed to the ideal incomplete system of the model can be found in a host of different physical systems exhibiting a particular pattern of behaviour, the incomplete system is embedded in its target systems and it therefore defines the shared universality class. The idea is that instead of separately studying each type of these systems, one instead studies the mathematically tractable minimal model and with the right modifications and adjustments derives conclusions about the behaviour of the physical systems. In other words, different physical systems can be categorised in virtue of a set of shared properties that are responsible for a particular pattern of behaviour –

---

[25] It should be noted here that the sense in which Batterman and Rice use the term 'universality class' for the purposes of their argument is not the strict technical sense in statistical mechanics as discussed, for instance, by Franklin (2018). In that sense, different physical systems belong to the same universality class insofar as they exhibit the same behaviour as they approach the critical point, i.e. insofar as the have the same set of critical exponents $\{\alpha, \beta, ...\}$ for several power laws.

e.g properties responsible for exhibiting oscillating movement, properties responsible for exhibiting two-body orbiting movement, properties responsible for electron scattering etc. This minimum shared set of properties is then attributed to an ideal incomplete system which in turn defines the universality class comprising all the physical systems in which it is embedded, providing this way a useful classification of nature.

1.4.2 Explanation

In addition to the neat description and classification of various physical phenomena, physics often provides *explanations*. In general, an explanation in physics can be understood as an answer to a 'why-question' regarding the existence or the behaviour of certain physical systems, the observation of regularities and repeatable patterns in nature, as well as the observation of certain patterns of data in experiments. A nice example from contemporary physics comes from the explanation of the flat rotation curves of galaxies in terms of dark matter. The rotation curves of galaxies are diagrams representing the orbital velocity of stars and gases in galaxies as a function of their distance to the galactic centre. Given that most of the stars of a galaxy are distributed around the galactic centre, the velocity of a rotating disk of stars and gas is expected to follow a Keplerian decrease ($u_r \propto r^{-\frac{1}{2}}$) in which the rotational velocity begins to decline beyond the radial distance containing most of the galaxy's mass. In the early 1970s, to the surprise of the scientific community, Vera Rubin and Kent Ford (1970) and Kenneth Freeman (1970) published two independent studies showing that, contra to our current gravitational theories, the rotation curves of the nearby observed galaxies tended to be flat, after an initial rise attributed to the central bulge (Fig.1.4). The 'why-question' to be explained here was 'Why the orbital velocity of galaxies does not decrease proportionally to their radial distance from the centre of the galaxy, according to our best theories of gravity'. Today, this question is answered by postulating a density distribution of stars according to which all galaxies are embedded in a large halo of dark matter which is responsible for the observed orbital velocities of stars in the outer regions of the galaxy. As we shall see in Chapter 5, whether this is a genuine explanation for the flatness of galactic rotational curves remains a moot point.

Contemporary discussions in philosophy of science concerning scientific explanation began with the development of the Deductive-Nomological (DN) theory of explanation by Carl Hempel (1965) and Hempel & Oppenheim (1948). Roughly speaking, in Hempel's account (1965, pp.247-51) explanations are answers to 'explanation-seeking why questions' that take the form of a deductive argument. In order for the explanation

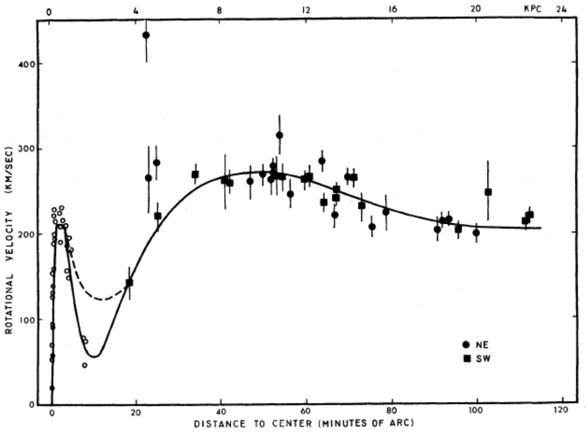

Figure 1.4: The rotation curve of the M31 galaxy (also known as the Andromeda nebula) as it appeared in the original paper by Vera Rubin and Kent Ford in 1970. After an initial rise in the 10-40 arc-minutes region, the orbital velocity of the members of the galaxy flattens, contrary to the declining prediction based on Kepler's laws.

to be successful, the premises of the argument must all be true and contain at least one natural law which is necessary for the derivation of the conclusion. A standard – and often used – example of a successful DN explanation is the answer to the question 'why the volume of a given gas increases when heated' which can be given in the form of a deductive argument as follows:

P1. The gas is kept at a constant pressure.

P2. The gas is heated up.

P3. The ideal gas law states that: PV=kT

Therefore, the volume of the gas increases.

Hempel's account, although greatly influential in the literature following its formulation, faces some well-known difficulties especially due to the famous problem of symmetry, first pointed out by Bromberger (1966), and the fact that it allows irrelevant premises to be added in a deductive argument that leads to a certain conclusion.[26] These objections sparked the interest of philosophers in scientific explanation and led

---

[26] See Salmon (1989, pp.46-50) for a discussion regarding a number of well-known counterexamples to the DN model. Ruben (2015, Ch.6) also provides a nice summary of the symmetry and the irrelevance objection to the DN model.

to the development of further theories of explanation that would – one way or another– eschew the difficulties that plagued the DN model.

The most notable examples of such work are perhaps Salmon's Statistical Relevant (1971), and Causal Mechanical (1984) theories of explanation, and Kitcher's (1989) Unificationist account of explanation.[27] These classical accounts of scientific explanation share the conception that science often provides explanations, and their common task is to identify the necessary and sufficient conditions for something to count as a successful scientific explanation of physical facts, as opposed to a mere description. What they do not do however, is to take into consideration the contribution of scientific models to the production of good scientific explanations. This task has been the focus of a more recent literature on model explanations which largely began with the work of Ernan McMullin (1978) on Hypothetico-Structural (HS) model explanations. For McMullin, the behaviour of various complex physical systems is explained by alluding to a *structure* of the system, where 'structure' is to be understood as the set of the constituent entities of the system and the relations between them. This structure 'is often called a physical (or a theoretical) model' (*ibid.*, p139). What makes structural explanations hypothetical is the fact that the structural model of the system 'is *postulated* to account for the observed properties or behaviour of the entity under investigation' (*ibid.*, p.139) and is therefore tentative, in that it is always possible that a different structure can also explain the features of the system. The explanation is achieved in virtue of the fact that the selected features of the model are causally responsible for the phenomenon to be explained; that is, they successfully capture the causes and their effects in the physical system to be explained.

What is of special interest, is that McMullin cites as one of the main reasons for the powerfulness of hypothetico-structural explanations their ability to reach beyond the empirical realm by postulating a structure for the phenomena under investigation. The example he uses is the explanation of the heat-expansion of iron in terms of a molecular structure and the accompanying theory of heat in which the effect of temperature on molecular structure can be calculated. This penetration to the 'invisible realm' is for McMullin 'the real triumph of contemporary scientific explanation' since it allows ontological claims regarding entities beyond our senses and affords an understanding 'of a hitherto hidden world of processes and structures both macroscopic and microscopic' (*ibid.*, p.145). What one finds in McMullin's analysis is therefore a role of scientific models as investigative tools in the unobservable world. The speculative modelling of

---

[27] See Woodward and Ross (2021) for a comprehensive review of these accounts.

unobservable entities not only provides possible explanations for complex observable phenomena, but it also allows the construction of possible theories about the behaviour of physical systems both in the macroscopic and the microscopic scales in which the perception of our senses and instruments is limited.

The causal character of model explanations is also reflected in Craver's (2006) account of mechanistic explanations. For Craver, scientific models are explanatory when they describe mechanisms, where mechanisms are understood as constituting of 'entities and activities organized such that they are productive of regular changes from start or set-up to finish or termination conditions' (Machamer et al. 2000, p.3). Following Woodward's (2005) manipulationist construal of causation, Craver argues that models that describe mechanisms are explanatory because not only they show how a system behaves, but they also say how it will behave under a wide range of interventions, and hence, they can be used to answer various "what-if-things-had-been-different" questions. This is opposed to 'merely phenomenal models' which, in principle, aim to provide an empirically adequate mathematical framework that is heuristically useful for accommodating results and deriving predictions while at the same time remaining silent about the internal structure of the target system and the underlying mechanisms that produce the results.

In order to provide an adequate explanation via a mechanistic model Craver lists four normative requirements that need to be fulfilled. The first is that the model must provide an accurate and complete characterisation of the phenomenon to be explained. This includes providing the precipitating, inhibiting, modulating and non-standard conditions of the phenomenon which altogether capture the multiple facets of the phenomenon to be explained. A mis-characterization or a partial characterization of a phenomenon can lead to a failure of explanation. The second requirement is the accurate description of the various parts of the mechanisms. In order to have a successful mechanistic explanation, the parameters and the relations between them in a mechanistic model must correspond to real component entities and activities, and not fictional posits. Real parts can be distinguished by their robustness – i.e. the fact that they are detectable with multiple causally and theoretically independent devices – and their manipulability – i.e. the fact that they can be used to intervene into other components and activities. The third requirement concerns the activities of mechanisms, that is, the things that the entities of a mechanism do. For Craver, the activities in mechanisms of successful explanations should not be understood as merely input-output pairs. Rather, they must be understood 'in terms of the ability to manipulate the value of one variable in the description of the mechanism by manipulating another' (*ibid.*, p.372). Thus, if two features of a mechanism X and Y (or parameters of a model) are connected to each other via an activity, then it should be possible to manipulate Y by manipulating X or vice versa. This requirement of manipulability serves as a criterion to distinguish causally relevant from causally irrelevant factors, which is extremely important for a successful causal explanation. Finally, the last crucial requirement for mechanistic explanations is the organization of the components of the mechanism in a way that exhibits the phenomenon to be explained. To provide mechanistic explanations is not merely to describe an aggregate of the properties of a system; one must also show how the different features of a phenomenon depend upon the organizational features of the underlying mechanisms, such that one cannot typically add or remove any parts in the mechanisms without disrupting the behaviour of the whole.

At first glance Craver's requirements might seem a bit stringent, especially when one takes into consideration that one of the most powerful characteristics of models is that they are mathematically tractable idealised versions of their target systems, and hence, they often fail to provide accurate and detailed descriptions of physical phenomena. Craver addresses this issue by pointing out a useful distinction between *ideally complete* and *pragmatically complete* models. The former are models that include all relevant features of a mechanism, while the latter only satisfy the pragmatic demands implicit in the context of the request for explanation. Hence, a mechanistic model is potentially a successful explanation insofar as it captures all the relevant features of a mechanism that are responsible for the production of a given phenomenon, regardless of whether it provides a full description of the physical system under investigation. This indicates once again the usefulness of thinking about the ideal systems in models as incompletely determined objects. By selectively attending only to the relevant features of a physical mechanism one builds a simplified ideal system whose mathematical tractability allows the derivation of results and the traceability of the causal processes of the mechanisms as they appear in the mathematical relationships of the mathematical framework of the model.

One may notice here an interesting tension between McMullin's Hypothetico-Deductive explanations and Craver's mechanistic explanations. For McMullin the fact that an explanation is hypothetical provides a gateway to the 'unobservable world' in that it paves the way for the expansion of scientific theories in microscopic and macroscopic scales, while for Craver, the 'fictional posits' of scientists in mechanistic models are heuristically useful but not explanatory. This tension arguably reflects the important distinction between how-possibly and how-actually explanations. This distinction has a long history that goes back to the writings of William Dray in the philosophy of history (Dray 1957), and its traces can also be found in Hempel's categorization of explanations as 'potential', 'more or less supported by evidence', and 'true' (Hempel 1965, p.338). In general, how-possibly explanations can be understood as speculative explanations indicating one or more possible scenarios for the production of a certain observable phenomenon, whereas how-actually explanations are explanations that capture the real nature of the phenomenon to be explained. Brandon (1990, p.184) nicely summarises the distinction between how-possibly explanations and how-actually explanations by stating that:

> A how-possibly explanation is one where one or more of the explanatory conditions are speculatively postulated. But if we gather more and more evidence for the postulated conditions, we can move the how-possibly explanation along the continuum until finally we count it as a how-actually explanation.

On Brandon's view the distinction between the two types of explanation is a matter of the degree of confirmation, and thus, the two kinds of explanation can be understood as lying on a continuum with respect to the empirical evidence supporting their claims. As we collect more evidence that the processes cited in a model correspond to the processes operating in nature, we gradually shift from a how-possibly to a how-actually explanation. The tension between McMullin's and Craver's accounts stems from the fact that for the former a *good* explanation need not necessarily be a *correct* explanation, while for the latter the correspondence of a model mechanism to a physical mechanism is a necessary condition for an adequate explanation.

Craver's interpretation of the distinction between how-possibly and how-actually explanations is that the former are 'loosely constrained conjectures about the mechanism that produces the explanandum phenomenon' as opposed to how-actually models that 'describe real components, activities, and organizational features of the mechanism that in fact produces the phenomenon' (*ibid.*, p.361). What Craver implies is that the distinction between how-possibly models and how-actually models lies in their intrinsic nature and not the degree of empirical confirmation. The former are, according to Craver, heuristically useful in constructing a space of possible mechanisms, but they are not adequate explanations, since for Craver, an adequate explanation must also be true.

In response to Craver, Bokulich (2014, p.334) follows Brandon and argues that the distinction between how-possibly models and how-actually models does not depend on the amount of detail that is present in the model, but rather on whether the represented model mechanism is indeed the mechanism operating in nature. As an example, Bokulich uses the case of tiger bush, a periodic banding of vegetation in semi-arid regions that forms stripes within a certain range of wavelength separated by barren areas. The explanation for the particular patterns of stripes in the tiger bush lies, according to Bokulich, in the fact that the relevant model in the simulation is specified in a rather abstract manner, and the more finely the explanatory mechanism is specified, the less confident scientists are that their detailed characterization is indeed the actual one found in nature. Bokulich thus illustrates a case in which a highly abstract model provides a genuine explanation for the stripes of tiger bush by virtue of its idealizations.

Bokulich's response stems from her own account of model-based explanations (Bokulich 2011) inspired by Woodward's counterfactual account of explanation, in which 'the explanation must enable us to see what sort of difference it would have made for the explanandum if the factors cited in the explanans had been different in various possible ways' (Woodward 2005, p.11). The core idea of Bokulich's account is that all appropriate model explanations must share the following three features: (i) the explanans makes exclusive reference to a scientific model that represents a specific target system or phenomenon to be explained, (ii) the model explains the explanandum by showing how the elements of the model correctly capture the patterns of counterfactual dependence in the target system, enabling this way one to answer various 'what-if-things-had-been-different' questions, as Woodward describes them, and (iii) there is an appropriate 'justificatory step' specifying the domain of applicability of the model and showing where and to what extent the model can be trusted as an adequate representation – and explanation – of the target system/phenomenon to be explained.

The take home message is that scientific models provide both how-actually and how-possibly explanations, regardless of the exact terms in which one understands this distinction. In general, how-possibly explanations can be thought of as speculative explanations for a given set of data that cannot be accommodated by an established theory, and thus, they are often – but not necessarily – linked with phenomenological models. Prime examples of such models are the various phenomenological models of dark matter to be discussed in Chapter 5, which are built on the basis of the available empirical data supporting the existence of dark matter but, nonetheless, do not fit in the well-established theory of the standard model. As a model gradually becomes part of a well-established theory and moves towards the theoretical model side of the spectrum, it benefits from the wider predictive success of the background theory, and the provided model-explanation becomes more and more plausible, shifting towards a how-actually explanation. Nevertheless, the importance of models in providing a scientific explanation, be it a how-actually or a how-possibly explanation, lies on the fact that often these explanations necessarily require the adjustment of the theory to a physical phenomenon, which as we have seen in the previous sections, can only be achieved via the construction of a representational model.

Finally, let us also note that regardless of which account of model explanation one endorses, the presented four-features framework provides a solid conceptual basis for understanding the role of models in explanations. In McMullin's Hypothetico-Structural account the crucial element for the explanation of a physical phenomenon is the *structure* of the system comprising the constituent entities of the system and the relations between them. This structure can be understood in this context as a simplified and incomplete ideal system, whose attributed properties are related according to the mathematical framework of the model and its interpretation. In a similar spirit, the crucial role of mechanisms in Craver's account can also be understood as stemming from the construction of an ideal incomplete system that is characterised only by those properties that are necessary for the function of the mechanisms. Craver's second requirement that the parameters and the relations between them in a model must correspond to real components of mechanisms, is reflected by the development of an appropriate interpretation of the mathematical framework of the model, according to which certain variables and parameters correspond to certain physical magnitudes that are responsible for the behaviour of the mechanism. Similarly, the causally relevant factors of a model – and subsequently of a physical system – are identified and distinguished from irrelevant factors via their treatment within the mathematical framework of the model.

1.4.3 Prediction

The third and final major achievement of physics is the derivation of results concerning future events, that take the form of predictions. The ability of theory to anticipate experiments was seen by Duhem as one of the major reasons for the strong belief of scientists that the theoretical classifications of nature provided by the theory are real. Among the infinitely many hypotheses that can be derived from a theory, says Duhem, we focus on a small number of theoretical consequences 'which do not correspond to any of the experimental laws previously known, and which simply represent possible experimental laws' (1954, p.28). Among these few consequences, the most interesting ones are those that are realisable in practice since these are the only hypotheses that can be submitted to test by facts. If the tests are successful the theory will be augmented. If not, the theory then needs to be 'more or less modified, or perhaps completely rejected' (*ibid.*).[28]

It is well known that this 'testability' of theories described by Duhem later characterised the work of Karl Popper, who was one of the most prominent advocates of the power of predictions in science. For Popper, the ability of scientific theories to make predictions is what distinguishes them from pseudoscience, since these predictions provide the grounds for severe tests that will falsify or corroborate an existing theory. Indeed, the confirmation of a novel theoretical prediction was for Popper the only sort of evidence that counts, since only those theories that survive such tests should be 'tentatively accepted' until falsified (1963). Imre Lakatos subsequently incorporated Popper's views on prediction in his theory of the scientific method. For Lakatos, scientific progress is defined in terms of novel predictions: a research programme is 'theoretically progressive' insofar as the theory behind it predicts a novel, unexpected fact, and 'empirically progressive' insofar as the predicted novel facts are established by experiment (1970, p.118). Competing research programmes are then evaluated by their empirical progress, since those programmes that are empirically more progressive will eventually defeat their rivals and thrive.

Lakatos' understanding of progress in science in terms of novel predictions sparked a philosophical interest in the concept of novelty. Both Popper and Lakatos endorsed a temporal condition of novelty by initially claiming that a prediction counts as novel only if it is not known to be true at the time it is derived by the theory. This view, however, was later challenged by Ellie Zahar (1973), who showed that if one insists in equating novelty with temporal novelty one ends up in a paradoxical situation in which several well-known predictions in science lose their significance. For instance, says Zahar, Einstein's derivation of the anomalous precession of Mercury's perihelion from General Relativity loses its significance since this fact was well known before the theory was proposed, and therefore, according to Lakatos' definition it is not a novel prediction. As a solution to this paradoxical situation, Zahar proposed a new criterion for novelty according to which 'a fact will be considered novel with respect to a given hypothesis if it did not belong to the problem-situation which governed the construction

---

[28] In Chapter 4, we shall investigate an example of this methodology in which one of the theoretical consequences of the standard model is submitted to test by experiment via its comparison to a data model.

of the hypothesis' (1973, p.103). Hence, the derivation of the exact amount of perihelion advance in Mercury from General Relativity still counts as a novel prediction, since it is was not included in the initial problem-situation for the construction of the theory. In other words, the theory of General Relativity was not constructed in order to explain the precession of Mercury's perihelion, but nonetheless, managed to accommodate this result and thus increase its credibility.

This short debate between Zahar and Lakatos uncovered the importance of the distinction between the *prediction* and the *accommodation* of results in science. A prediction is largely understood as an empirical consequence of a theory that has not yet been observed at the time the theory is constructed, whereas an accommodation is an empirical consequence of a theory that is known at the time of the construction of the theory, but which, nonetheless, does not belong to the initial set of observations that contribute to the construction of the theory. Hence, Zahar's criterion of novelty can be understood as a broadening of the concept of novel facts to include instances where theories accommodate previously known facts that are not involved in the construction in the theory. The question that deeply concerned philosophers of science was whether the prediction of novel facts from a theory lends stronger support to the theory than the accommodation of known facts. A large part of the relevant literature that followed Lakatos' remarks on novel facts and predictions led to the development of various strands of *predictionism* (or predictivism), the philosophical view that the prediction of empirical facts by theories is intrinsically superior to the accommodation of previously known facts.

The details of the debate on the prediction and accommodation of results by theories and whether prediction is indeed superior than accommodation, lay outside the scope of this chapter and need not concern us further here.[29] Rather, what we are mostly interested in is the role of models in the derivation of the hypotheses that lead both to predictions and accommodations of results in physics. The first thing to note is that the derivation of concrete empirical consequences from a theory that can either predict novel facts or accommodate existing phenomena is a process that necessarily proceeds only via the construction of representational models, that serve as the connecting bridge between the abstract worlds of theory and mathematics and the concrete nature of physical phenomena. This is perhaps a trivial point, but it, nonetheless, highlights the often neglected fact that the application of physical theories to real phenomena is a process that necessarily requires skilful and successful modelling.

---

[29] For a comprehensive review of the literature on prediction and accommodation, see Barnes (2021) and references therein.

The second thing to note, is the close connection of phenomenological models with the accommodation of results. As already mentioned, by definition, phenomenological models are constructed in order to accommodate empirical phenomena that the existing theoretical models cannot. For instance, Bohr's model of the hydrogen atom, was constructed to accommodate – among other things – the discrete nature of the spectral lines of hydrogen, a known fact that Rutherford's theoretical model failed to explain. This is not to say that theoretical models can only give predictions however. (The application of general relativity to the perihelion of Mercury is, in fact, an example of how theoretical models can also be used for the accommodation of known facts by the theory). Nor does this mean that phenomenological models are only constructed to accommodate known facts and cannot give any predictions. On the contrary, instances where a phenomenological model that was initially constructed to accommodate a known fact succeeds in making novel predictions of other not yet observed phenomena are not rare in physics, and are often seen as a major success that initiates radical revisions in existing theories.

A nice example of a famous prediction in physics, is the prediction of the anomalous magnetic moment of the electron by Julian Schwinger in 1948. The anomalous magnetic moment of the electron represents the effects of quantum electrodynamics to the classical magnetic moment of electrons, which can be thought of as a measure of the strength of the magnetic field of the electron. The classical magnetic moment of the electron μ_0 is obtained from the Dirac equation in terms of the g-factor, a dimensionless quantity that characterises the magnetic moment and the angular momentum of a particle, and for which the Dirac equation gives $g = 2$. The anomalous magnetic moment of the electron α_e is the difference between the Dirac magnetic moment and the quantum electrodynamics result defined by

$$\alpha_e \equiv \frac{g-2}{2} \approx 0.0011614 \tag{1.17}$$

Schwinger's (1948) perturbative analysis in quantum electrodynamics showed that the classical value derived from the Dirac equation only corresponds to the first (tree-level) Feynman diagram for electron scatterings and does not take into account the contribution of higher order terms. His major achievement was the successful calculation of the first (one-loop) correction of the magnetic moment of the electron, which he found to be

$$\delta_\mu = (\frac{\alpha}{2\pi})\mu_0 \tag{1.18}$$

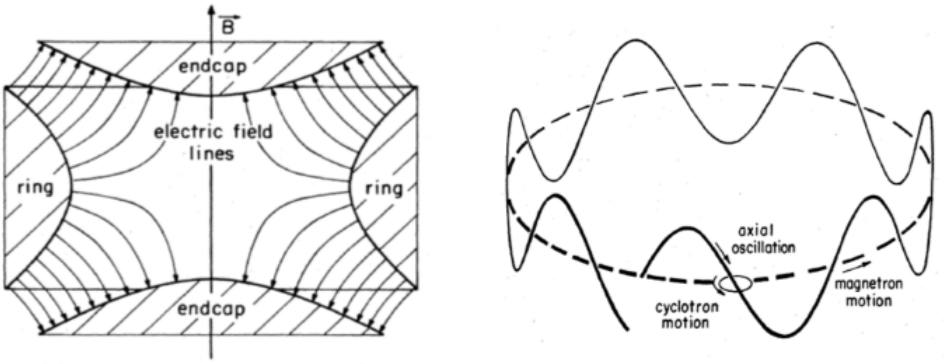

Figure 1.5: Electric and magnetic field configurations of the Penning trap (left) and orbit of a charged particle in a Penning trap (right), as found in Brown & Gabrielse (1986).

where α is the fine structure constant ($\approx 1/137$). Schwinger's result essentially accommodated a previously known experimental result (Kusch & Foley 1947) indicating a small deviation between the classical magnetic moment and the observed value of the magnetic moment of the electron, but it is interesting to note that it is widely referred to as one of the most significant *predictions* of QED. Today, the anomalous magnetic moment of the electron has been calculated up to the tenth-order perturbation theory, which corresponds to the corrections up to the fifth power of α^5 (Aoyama et al. 2018). The current theoretical value of α_e agrees with experimental results to more than ten significant figures, making the magnetic moment of the electron the most accurately verified prediction in the history of physics.

Schwinger's seminal work on the perturbative analysis of QED and the subsequent efforts to derive the theoretical value of the anomalous magnetic moment of the electron via the calculation of thousands of Feynman diagrams are essentially attempts to derive the empirical consequences of the theory of quantum electrodynamics. This process involves the development of the abstract mathematical framework of the theory into a mathematical framework of a representational model for a concrete physical phenomenon, which with an appropriate interpretation can be linked to an experiment. In the case of the anomalous magnetic moment of electrons, the experimental measurement of this quantity typically involves the use of a Penning trap.

Penning traps are devices that can capture single charged particles by combining a homogeneous axial magnetic field and an inhomogeneous quadrupole electrostatic potential, allowing the precise measurement of various properties of subatomic particles.[30]

---

[30] Penning trapped particles are often called 'geonium atoms' since their only binding is to an external

Ultimately, the connection between theory and experiment is achieved with the construction of a representational model for the physics of single electrons in Penning traps (Brown & Gabrielse 1986). The main idea behind the measurement of the anomalous magnetic moment of the electron in a Penning trap is that the difference between the electron's cyclotron frequency – i.e. the frequency of the electron moving perpendicular to the direction of the magnetic field – and its spin precession frequency – i.e. the frequency in which the angular momentum vector precesses about the magnetic field axis – is proportional to $g-2$ which is closely associated with the magnetic moment of the electron.

To use the terminology of Sec.1.2, the mathematical framework of the model of the single electron in a Penning trap essentially provides the mathematical relations between all the necessary quantities for the measurement of the various properties of single electrons, including the relation between the electron's g-factor and the difference between the electron's cyclotron frequency and its spin precession frequency. The interpretation of this framework links the mathematical objects of the model to measurable physical quantities, and gives rise to an ideal system that resembles a 'geonium atom', i.e. a single electron in a Penning trap under the influence of a magnetic and electrostatic potential. These features are expressed and communicated via various representational means, such as the two diagrams in Fig.(1.5). Schwinger's perturbative analysis for the theoretical derivation of the anomalous magnetic moment of the electron can be seen as the intermediary step between the development of the mathematical framework of the theory to the mathematical framework of the model for an electron in the penning trap. Firstly, one develops the general principles of quantum electrodynamics in order to derive – with the use of approximating techniques such as perturbation theory – the empirical consequences of the theory, and only when this first step is accomplished, the empirical consequences are adjusted to the refined data – i.e. the data model – of a concrete physical experiment via the construction of an empirically relevant model. As we shall see in Chapter 4, what is eventually compared with the data model is the context specific result of the representational model whose interpretation gives rise to an ideal system that resembles the target system of the experiment.

---

apparatus residing on Earth. For a review of the experimental capabilities of Penning traps see Blaum et al. (2010).

1.5 Conclusions

The analysis of the three major achievements of physics and the role of models therein brings us to the end of the first chapter. Our discussion began in Sec. 1.1 with a brief review of the literature on the structure of scientific theories which essentially generated the vast literature on the philosophy of models that followed until present days. The first rigorous philosophical accounts on the structure of scientific theories by the Logical Positivists in the 1950's, resulted in the development of the Syntactic View which understood scientific theories as axiomatised collections of sentences in higher order predicate logic. However, the important limitations in the aspiration of the Logical Positivists to reconstruct scientific theories in the formal language of first order logic, naturally led to the development of the Semantic View, mainly by Suppes and van Fraassen. By viewing theories as sets of models instead of formal sentences, the Semantic View shifted the focus from the contents of theories to the contents of the scientific models that can be derived from a theory, and essentially initiated the deep philosophical interest in the nature of scientific models and their role in science. In the more recent literature, the tendency to reconstruct scientific theories in the formal languages of mathematics and logic by the Syntactic and the Semantic view was gradually abandoned in favour of more pragmatic approaches that highlighted the importance of various non-formal elements in theories, such as the continuity of theories and experiment, the plurality of models in science and other sociological aspects.

In the spirit of this pragmatic approach, the following section provided a useful conceptual framework for understanding the nature of scientific models as epistemological tools, in terms of four main features: the mathematical framework of the model, its interpretation, an ideal system, and the representational media by which models are expressed and communicated. This framework helps in distinguishing models from theories and sets the basis for the analysis of models throughout this thesis. In what follows, models will be understood as *epistemological tools* in science that are characterised by the aforementioned four features, and whose main aim is to provide knowledge about the behaviour and the properties of their target systems. The deconstruction of models in terms of their four main features is a useful way to understand how models act as a mediating vehicle between theories and physical phenomena, by showing how each feature of the model contributes to the specific aims of the modeller and the broader aims of physics in general.

We have also highlighted the important distinction between theoretical and phenomenological models in physics. The former are models that can be deductively derived from the fundamental principles of the background theory, modulo some controllable idealisations – i.e. idealisations and approximations that can be justified theoretically or mathematically. The latter are models whose construction typically begins with the aim of accommodating experimental data that cannot be accommodated by the existing theoretical models, and as a result, their mathematical framework often contains expressions that are incompatible with the fundamental principles of the relevant theory. The important thing to keep in mind, is that this distinction does not reflect a difference in the intrinsic nature of the model, nor is it a measure of the model's accurate representation of its target. Rather, the distinction reflects the different starting points for the construction of these two different types of models, as well as their different aims. Theoretical models are generated from the theory and their aim is to test and validate the empirical consequences of the theory, while phenomenological models are generated from the data and their aim is the accommodation of unexpected data, which will eventually act as the starting point for the refinement of the theory.

In Chapters 3, 4, and 5 we shall see how these two types of models play different – yet crucial – roles in the methodology of physics. In particular, in Chapter 3 we will explore the special case of models in perturbative quantum field theory that seems to indicate the existence of a third special type of models that are neither theoretical nor phenomenological, and argue that the presence of such models in one of the most successful physical theories poses a significant challenge to van Fraassen's constructive empiricism and the Semantic View in general. In Chapter 4, we shall see how a theory is put to the test via the comparison of a theoretical model with a data model, by closely examining an example of the experimental tests of Lepton Flavour Universality at the Large Hadron Collider. The discussion in this chapter nicely illustrates how the deviation between the theoretical models and the data models gives rise to the development of new phenomenological models, and highlights the importance of various methodological considerations in the construction of data models. Finally, in Chapter 5, we shall explore the methodology of dark matter observation and the role of the various phenomenological models of dark matter therein. As we shall see, the extrapolation and validation of results that will eventually provide new insights about the nature of dark matter faces important limitations.

Before we delve deeper into these methodological issues however, in the next chapter we shall defend the pragmatic approach to the ontology of models that was presented in this chapter. Recall that in Sec. 1.2 we pointed out that the presented account of models is merely a useful *conceptual framework* for the understanding of models as epistemological tools in physics and that our claims should not be seen as making any sort of ontological commitments about the existence of abstract incomplete objects in scientific models. It was also noted that the four main features by which models are characterised are not necessary and sufficient conditions for something to count as a model and thus, they should not be seen as the defining features of models. Nonetheless, by providing a characterisation of the nature of models in terms of these four features, we have engaged in an attempt to answer a question that has been puzzling philosophers of science for several decades. That is, we have attempted to answer the question ‘What is a scientific model?’, a question that lies at the heart of a long-standing debate on the ontology of models. The next chapter provides a critical analysis of the literature on this issue and offers a possible solution in Carnapian terms that avoids the various metaphysical conundrums and other difficulties in this debate that otherwise seem insuperable.

CHAPTER 2

A PRAGMATIC APPROACH TO THE ONTOLOGY OF MODELS

As we have seen in the first chapter, scientists use models for a number of different purposes and these models come in a variety of forms. A natural question to ask is therefore

> [Q] What are scientific models?

A long-standing debate in the literature on scientific models concerns the possible answers to this question. While some authors have voiced their scepticism that this question has a meaningful answer (Callender & Cohen 2006; Suárez 2004; French 2010), others have tried to give a more positive note by arguing that models are best understood as real existing abstract objects (Giere 1988; Psillos 2011), fictional entities (Godfrey-Smith 2006; Frigg 2010; Toon 2010, 2012) and mathematical structures (van Fraassen 1980; Da Costa & French 2003). Each of these accounts comes with its own strengths and weaknesses and faces its own difficulties in giving a conclusive answer to the main question on the ontology of models.

The primary aim of this chapter is to adopt a Carnapian meta-ontological stance and show that some of these ostensibly insurmountable difficulties stem from an inappropriate reading of [Q] as a purely metaphysical question. Building on Carnap's (1950/2012) tripartite distinction between (i) internal questions (ii) external practical questions and (iii) external theoretical questions, it will be shown that [Q] should be understood as an *internal theoretical* question within an already accepted linguistic framework or an *external practical* question regarding the choice of the most appropriate form of language in order to describe and explain the practice of scientific modelling. The further reading of [Q] as an *external theoretical* question, that is, as a question about the *real* nature of models, independently of any form of language that might be used to describe them, is deceptive and should be avoided. The conclusion is that, from a Carnapian perspective, the debate on the ontology of models is ultimately about the choice of an appropriate language in order to describe the practice of scientific modelling, and as such, it does not admit of a *unique true answer*. By adopting different 'ontologies' of models, philosophers are in effect advocating for the various alternative ways by which one can understand the abstract nature of models and their role in scientific inquiry.

Rather than arguing for the supremacy of a Carnapian meta-ontological stance in general, the aim here is to take the Carnapian framework as a working premise and demonstrate the implications and payoffs of this view on the debate about the ontology of scientific models.[1] It should also be noted that if one is sceptical about the Carnapian programme in general, there is nothing special in the debate about the ontology of models which favours the adoption of a Carnapian stance specifically about this matter. The aim is therefore not to argue that one should be a Carnapian with respect to the ontology of models, regardless of one's beliefs in other issues of metaphysics. Rather, the aim is to illustrate that once one adopts the Carnapian perspective, a number of issues in the debate are resolved and the focus can be shifted to other non-trivial questions about the general practice of scientific modelling which will indeed concern us for the rest of this thesis. Carnap's motivation in applying his method to metaphysics was to bring to philosophy the kind of progress that is usually found in the natural sciences, and this chapter aims in showing how this progress can be achieved in the debate on the ontology of models by applying the Carnapian method.

More specifically, the proposed understanding of the debate in Carnapian terms teaches us that the choice of an appropriate linguistic framework – i.e. the choice of an appropriate ontology – is only a practical matter relative to the aims for which the

---

[1] Whether one has good reasons to adopt a neo-Carnapian stance in metaphysics in general, is something that has been discussed extensively in the relevant literature and the reader is referred to the original works of Carnap (Carnap, 1937, 1996, 2012; Carnap & Schilpp, 1963) and Quine (1951a, 1951b, 1960) for more detailed arguments and responses. In the more recent literature on meta-metaphysics, a number of compelling arguments towards a neo-Carnapian point of view can be found in the works of Huw Price (Price 2004, 2007, 2009; Macarthur and Price 2007) and Amie Thomasson (2014). A number of responses on the basis of neo-Quinean concerns can be found in the works of Sider (2009), Finocchiaro (2019), and van Inwagen (2020).

language is introduced. Hence, given that the aim of philosophical investigations on the nature of scientific models is to understand as much as possible about their function as epistemological tools in science, the various existing accounts should not be seen as competing and mutually exclusive theories aiming to find a unique true answer to the question of the ontology of models. Rather, they should be seen as complementary accounts that enable us to understand the different aspects of modelling. The main implication of this view is that the question of the ontology of models is only taken as a means of probing other related questions regarding the methodology of scientific modelling, such as questions on the capacity of models to provide knowledge, their relationship with background theories and their contribution towards the fulfilment of the main aims of science discussed in the previous chapter.

Indeed, the four-features account given in the introductory chapter of this thesis and the claim that (some) models comprise ideal systems that can be described in Meinongian terms, is not presented here as the unique and true account of the ontology of models, nor does it imply any sort of ontological commitment to abstract incomplete objects. Rather, in the spirit of a Carnapian pragmatic approach, it is, as already emphasised several times in the text, a useful conceptual framework that can be used as a basis for answering further philosophical questions about the methodology of modern physics and the central role of models in therein. The present chapter essentially motivates this Carnapian approach by highlighting the benefits of adopting a pragmatic stance on the question of ontology of models and viewing it as an intermediate step within the overall project of this thesis, which is the evaluation of the role of models in the methodology of modern physics.

The realization that the debate on the ontology of models can be reframed in Carnapian terms effectively dissolves the debate and urges philosophers to move forward, by arguing that there is nothing more to be gained in trying to settle on a unique true answer to the question of the ontology of models. The main argument is that the two proposed readings of [Q] jointly provide all the necessary conceptual tools for developing a robust theory of models, whilst keeping away from the various insuperable challenges faced by the aforementioned existing accounts. The onus is thus on the proponents of such views in the sense that they need to show what the extra benefit of attempting to settle on a conclusive answer is.

The chapter is organised as follows. In Section 2.1 the problem of the ontology of models is described in more detail with references to the relevant literature. In Section 2.2, Carnap's distinction between internal and external questions is presented, followed by a discussion of how this distinction can be exploited for the development of a theory of models, and an analysis of the difference between the Carnapian approach and other possible ways of rejecting metaphysics. Section 2.3 discusses French's main argument for quietism as a possible route towards a pragmatic approach. The argument is found susceptible to a number of objections and thus further justification is needed. Finally, in Section 2.4, an objection to the proposed pragmatic approach is addressed. The conclusion is that the objection does not succeed in rendering pragmatism about the ontology of models an unattractive position.

2.1 The problem of the ontology of models

More than thirty years ago, Giere (1988) presented a theory of models as abstract systems that possess the properties ascribed to them and satisfy the equations by which they are governed. Giere's theory has been highly influential in the large discussion that followed regarding the ontology of scientific models, and which still carries on unresolved. By separating models from their descriptions, Giere ascribed to the former a status of independent abstract entities for which certain ontological questions regarding their existence and nature should be answered. With this in mind, Thomson-Jones (2010) has more recently described scientific models as 'missing systems'. These missing systems have the surface appearance of a precise description of actual concrete objects, however, we know that there are no such objects in the actual world fitting that description. The challenge is therefore to find an appropriate way to understand the nature of scientific models as missing systems, and it is often referred to as 'the problem of the ontology of models'.

A further motivation for tackling the question of the ontology of models is the fact that it is closely connected to the puzzling question of scientific representation; that is, the question of the exact nature of the relationship between models and the physical systems they represent, which Giere described in terms of similarity. The standard argument is that if representation is a relation between models and physical systems, and if models indeed carry some kind of representational capacity, then the only way to flesh out the nature of this relation is by providing a detailed ontology of models.

It is no surprise then that several attempts have been made so far to provide a positive account on the ontology of models. For example, following Giere, Psillos (2011) takes models to be real existing abstract objects, whereas authors like Godfrey-Smith (2006), Frigg (2010) and Toon (2010, 2012) have argued that models are useful fictions which, literally speaking, do not exist. An alternative approach, stemming from the seminal works of Suppes (1960) and van Fraassen (1980) on the semantic view of theories, focuses on the mathematical aspect of models and sees them as mathematical structures that represent physical targets in terms of some form of isomorphism. Da Costa and French (2003) are also strong opponents of this approach.[2]

This ongoing reflection on the problem of the ontology of models during the last three decades has, unsurprisingly, led to a further discussion regarding the metaphysics of abstract objects and their properties, bringing forward a host of difficult and well-known problems in traditional metaphysics. A standard objection against the abstract objects view – which could also target the description given in Chapter 1 in terms of incompletely determined objects – concerns the attribution of physical properties to abstract systems (Teller, 2001, p.399; Thomson-Jones, 2010, p.290). If models are existing abstract objects with no spatiotemporal location, then how is it possible for them to instantiate the spatiotemporal properties that make them similar to their targets? Similarly, against the fictionalist view the objection is that it is hard to see how a non-existing entity stands as a representation of a physical system in a way that allows a fruitful comparison between the two (Morrison, 2015, p.89). As for the structuralist approach, a standard worry is that if models are mathematical structures, then it is hard to understand how they stand in isomorphic relations with real systems (Frigg & Nguyen, 2017, p.71). What does it mean for a physical system to possess a structure, and where in that system is the structure located? These and other criticisms along these lines often come forward as challenges for all three main accounts on the ontology of models making the problem of ontology seem unresolvable.

French (2010) was the first to clearly point out the futility of trying to give a conclusive answer to the question of the ontology of models. By putting forward a quietist approach, French claimed that when it comes to questions about the *real* ontology of models and theories one should remain silent. Such a quietist conclusion musters support from the fact that metaphysical questions about the ontology of models and theories are both unanswerable and unnecessary, given that our aim is to understand and explain the scientific practice. This gives rise to the following two questions that need to be addressed:

[1] Can we answer questions about the ontology of models?

---

[2] For a recent review of the literature on the ontology of models see Gelfert (2017). See also Frigg and Nguyen (2017) for a review of model-based theories of representation.

[2] Do we need to answer questions about the ontology of models in order to understand and explain the practice of scientific modelling?

Notice how these two questions are connected with the central question of the ontology of models. The first question asks whether or not [Q] is an answerable question, while the second asks whether or not answering [Q] is necessary in order to have a fruitful theory of models. In what follows, it will be shown that the answer to these questions depends on whether [Q] is understood as an internal or external question in a Carnapian sense. As will become clear in Section 2.2, an understanding of [Q] as either an internal question or an external practical question trivialises both [1] and [2] and allows for a positive answer. This approach takes all references to abstract entities merely as a fruitful and efficient way of talking about scientific models and stays away from any form of metaphysical enquiry on the nature and existence of abstract entities. On the other hand, if [Q] is understood as an external theoretical question regarding the *real* ontological status of models, it then becomes a pseudo-question and therefore the answer to these questions is negative.

2.2 Unfolding the pragmatic approach

2.2.1 Internal and External questions

Carnap's principal goal in 'Empiricism, Semantics and Ontology' (1950/2012) is to clarify an ongoing bewilderment deriving from, what he calls, the problem of abstract entities. That is, the problem of referring to abstract entities, such as properties, classes, relations, numbers, propositions etc., while at the same time remaining faithful to the basic principles of empiricism and avoiding any sort of commitment to a metaphysical ontology of a Platonic nature.[3] What Carnap aims to show is that accepting a linguistic framework which involves reference to these abstract entities, does not imply the acceptance of the reality or existence of these entities in the traditional metaphysical sense, as understood, for instance, in the context of Platonism in mathematics. To be a Platonist about mathematical entities is to hold the view that abstract mathematical

---

[3] What is interesting here is that although Carnap explicitly states in the very first paragraph of his text (2012, p.241) that his focus is on abstract entities like numbers, properties etc., his overall approach is a general one against ontology and his distinctions essentially apply to all kinds of existence questions, including existence questions about physical entities such as electrons, black holes and so on. This point becomes clear later on through Carnap's thing-language example which will be discussed in the following paragraphs.

objects exist independently of us and our language. Carnap's claim is that existence claims about such entities are only meaningful *within* a linguistic framework.

A linguistic framework is a system of possible ways of speaking about new kinds of entities, subject to certain rules. In other words, it is a set of rules dictating the use of certain terms and predicates referring to new entities, such as properties and numbers, in order to be able to speak meaningfully about a given subject. Given a linguistic framework, Carnap makes a distinction between three types of questions concerning the existence or reality of the introduced abstract entities: (i) *internal* questions, (ii) *external practical* questions and (iii) *external theoretical* questions. As we shall see, for Carnap, the first two types of questions are legitimate and often trivial, whereas the latter is problematic. This may come as a surprise to those who read Carnap as rejecting all external questions, but as will be shown below, this is not the case. Carnap does welcome external statements about the existence of abstract entities, insofar as they are understood in a practical and pragmatic fashion.

Internal questions are questions asked *within a linguistic framework* and for which the answer can be found either by logical analysis or empirical observation. For instance, to use Carnap's example (2012, pp.244-5), once we accept the linguistic framework for numbers, the question 'Is there a prime number greater than a hundred?' is an internal question and the answer can be found by analysing the rules of the linguistic framework for numbers. In other words, in order to answer this question one merely has to check whether or not the existence of such a number follows from the rules of the already accepted system of numbers within which the question is raised. Questions like 'Is there a piece of paper on my desk' are internal questions within the framework of the 'thing-language' – i.e. the linguistic framework we use to speak about the external world – and the answer to such questions is a matter of empirical observation, since the rules of our chosen thing-language imply that a physical object exists if it can be empirically observed. Both logical and empirical internal questions are thus subject to the internal rules of the relevant linguistic framework. Internal questions are therefore *theoretical*; that is, they are questions for which there is a definite answer that follows logically or empirically from the rules of the relevant framework.

Internal questions are often (but not always) trivial, in the sense that a positive answer says nothing more than that the given linguistic framework is not empty. For instance, the question whether there exists a real number between five and six is trivial, since the answer comes easily from the rules of the linguistic framework for real numbers. Examples of less trivial internal questions are questions whether 'glueballs' exist or whether there is a prime number between nine billion and nine billion and ten for instance.[4] What makes an internal question non-trivial is the fact that the empirical observations about the existence of an entity may not be so clear – e.g. in the case of glueballs – or the fact that the application of the internal rules of the chosen linguistic framework may require extensive computational analysis – e.g. in the case of very large prime numbers.

External theoretical questions on the other hand, concern the existence of the system of abstract entities as a whole, *prior* to the acceptance of a new linguistic framework. Such questions are not raised within the scientific community or in common parlance, rather they are typically asked by philosophers in traditional metaphysics when, for instance, they pose the metaphysical question of the existence of natural numbers or the reality of the external world. What philosophers usually mean when they raise these questions is 'whether or not numbers [for instance] have a certain metaphysical characteristic called reality [...] or subsistence or status of independent entities' (*ibid.*, p.245). These ontological questions must be raised and answered, according to this approach, before the introduction of the new language. Hence, questions like 'Do numbers *really* exist?' or 'Is the external world *real*?' are external to the linguistic framework since the answer to these questions is supposedly independent of the language we use to speak about numbers and material things.

The problem with such external theoretical questions, Carnap says, is that they are devoid of any cognitive content; they are pseudo-questions. That is, they are ill-formed questions in the sense that they are '*disguised* in the form of a theoretical question while in fact [they are] non-theoretical' (*ibid.*, p.245, emphasis added). These disguised external questions cannot be answered, simply because it is impossible to frame them in terms of the common scientific language in a way which succeeds in giving them any cognitive content. To see why, recall that accepting a certain linguistic framework amounts to accepting a set of statements regarding the existence and nature of the abstract entities in question. For instance, within the system of numbers, the assertion that there is a prime number larger than one hundred simply states that this prime number is an element of the already accepted linguistic framework. However, the further external question of whether such a number *really* exists, is not part of the set of the

---

[4] In particle physics, glueballs are hypothetical colourless particles that consist only of interacting gluons without any valence quarks. The existence of glueballs is predicted by Quantum Chromodynamics but the results of various indirect experimental observations are still not universally accepted. For a relatively non-technical review of the physics of glueballs and their connection with the MIT bag model, see Mathieu et al. (2009).

accepted statements since it cannot be formulated in a meaningful way within this framework or any other theoretical language. In other words, the concept of existence cannot be applied to the system itself independently and prior to the acceptance of a given framework. As an alleged opponent of Carnap on this matter pointed out in a rather astute way, 'to ask what reality is *really* like [...] apart from human categories, is self-stultifying. It is like asking how long the Nile really is, apart from parochial matters of miles or meters' (Quine, 1992, p.9).[5]

Another useful way to understand Carnap's view on external statements is to compare them with moral statements under the scope of the more familiar doctrine of non-cognitivism in Ethics. For the non-cognitivists, moral statements such as 'Killing is evil' do not have any propositional content and thereby do not have any truth conditions. Rather, they only express beliefs and other non-cognitive attitudes such as revulsion and disapproval.[6] As one of the first non-cognitivists, Carnap also drew an analogy between metaphysical and ethical claims in his earlier works (1935/1996, pp.22-30; 1937, p.278) stating that the latter are mere commands in a misleading grammatical form, and thus they should not be treated as assertions. Similarly, metaphysical statements related to external questions – e.g. that numbers *really* exist independently of the adopted linguistic framework – only have an expressive function in that they only express personal beliefs. Nonetheless, they have no theoretical content and thus they should not be treated as truth-apt assertions.

This is not the end of the story however. External questions are indeed non-cognitive but this does not mean that they should be thrown out of the window. Rather, Carnap's insightful remark is that such questions should be understood as *practical* questions concerning the choice of a linguistic framework over another and the structure of rules within them. In other words, external questions like 'Do numbers *really* exist?' are questions concerning whether or not we should accept a linguistic framework with reference to numbers. However, the acceptance of a given framework, which further implies the acceptance of a set of (internal) statements regarding the existence of new entities, cannot be judged as being true or false simply because it does not involve an assertion.

---

[5] Contrary to the seemingly widespread view among philosophers which sees Quine as saving metaphysics from Carnap, this quote from Quine goes on to suggest that Quine's views on metaphysics are, to a large extent, on par with Carnap's meta-ontological stance. Price (2007, 2009) and more recently Verhaegh (2017) provide a convincing line of arguments to this direction showing that not only Quine does not undermine Carnap's main thesis, but in addition he 'overtakes him, and pushes further in the same direction' (Price, 2007, p.393).

[6] See van Roojen (2018), Blackburn (2006), and Schroeder (2010, esp. Ch.2) for more on Moral Noncognitivism.

Rather, it is a matter of a decision guided merely by pragmatic criteria such as the efficiency, fruitfulness and simplicity of the new language and the degree in which these new ways of speaking are conducive to the purposes for which the language was initially introduced. Nonetheless, the fact that a given language, such as the numbers-language for instance, turns out to be extremely efficient does not provide any sort of confirming evidence for the reality of numbers in the traditional metaphysical sense.

Before moving to the next section, it is important to stress the difference between internal theoretical questions and external practical questions. For Carnap, these are the only two legitimate ways to read existence related questions. The former admit of definite answers depending on the rules of the framework in which they are expressed, and thus, any internal assertion needs to be justified either by empirical evidence or logical analysis. External practical questions on the other hand, are questions of degree, and just like any other practical question, they do not admit of a definite answer. Rather, the answer to these kind of questions depends on pragmatic criteria relative to the purposes for which a linguistic framework is used. The further reading of external questions as theoretical questions for which a definite answer must be given stems from the fact that external questions are usually grammatically disguised as internal theoretical questions. However, this reading is problematic and should be avoided.[7]

2.2.2 Theories of ontology as competing frameworks

Carnap's conclusion is that the problem of abstract entities is a result of a failure to acknowledge this fundamental distinction between internal and external questions and I want to argue that the same holds for a large part of the debate around the ontology of models. The nature of the objections discussed in Section 2.1 and the fact that the debate appears to be unresolved show that [Q] is sometimes treated in the relevant literature as an external theoretical question for which there is a definite answer. However, from a Carnapian point of view, this reading is problematic and only succeeds in making [Q] an unintelligible pseudo-question. The suggestion here is that [Q] should be seen either as an internal theoretical question or an external practical question. The central question of the ontology of models is thus ambiguous and as we shall see, both readings are legitimate and serve different purposes. On the contrary, the further reading of [Q] as an

---

[7] For further contemporary discussions on the distinction between internal and external questions, as well as on the debate between Carnap and Quine on metaphysics see Bird (1995), Yablo (1998), Alspector-Kelly (2001), Eklund (2013), Verhaegh (2017, 2018), Morris (2018) and Flocke (2020). Blatti and Lapointe (2016) is a comprehensive collection of essays on Carnap's overall approach on ontology and metaphysics.

external theoretical question does not seem to improve our understanding of scientific models in a fruitful way and is therefore unnecessary. Let us elaborate on each one of these three options, beginning with the second.

As formulated above, [Q] is a question about which kind of abstract entities is to be identified with scientific models. As such, it can be understood as an external practical question asking: ‘What is the most appropriate and efficient form of language to describe scientific models?’. Given that this is a practical question, it only depends on pragmatic criteria and admits of multiple ‘equally true’ answers. The preference for a particular linguistic framework in the case of scientific models therefore depends on the specific desiderata for choosing an ontology of models over another. For instance, in our description of the ideal systems in models in the previous chapter, we have deliberately chosen Meinong’s language in terms of incompletely determined objects for its advantages in conceptualising certain aspects of scientific modelling. In other words, the choice for this framework was based on purely pragmatic grounds since the incompleteness of these objects helps to explain how the simplicity of models is crucial for the understanding and the representation of more complicated physical systems. Similarly, philosophers like Giere and Psillos opt for an abstract-objects-language (albeit with some differences) because they are primarily interested in explaining the attribution of physical properties like mass and momentum to highly idealized ‘non-existing’ systems, such as the model of a particle in a one-dimensional box in quantum mechanics. For Giere, an extra motivation for choosing an abstract-objects ontology is the development of a theory of representation in terms of similarity, whereby models and their targets share some of their properties. On the other hand, van Fraassen’s state space approach focuses on the mathematical nature of models and aims in capturing the ability of the latter to represent the evolution of the states of physical systems in time by the abstract nature of mathematical state spaces. Different accounts thus serve different desiderata and complement each other in that they offer different insights on the nature of models.

All of these views are entirely legitimate, and none of them should be judged as true or false simply because they should not be seen as assertions about the *real* ontology of models. The various accounts on the ontology of models should only be seen as representing different linguistic frameworks for speaking about scientific models in a fruitful and efficient way. The only meaningful comparison between them is therefore with regard to their success in being conducive to the aim for which they were initially introduced; namely, the aim of explaining as much of scientific talk about models as possible. What this means may vary from case to case and ultimately depends on the desired explananda of each account. Nevertheless, the ultimate aim should not be a definite answer to [Q], but the understanding of various model related questions such as how scientists build and use scientific models in different disciplines, what makes a model a good or bad epistemological tool for acquiring knowledge about a physical system, why scientists often use inconsistent models to represent the same physical system, what the relationship of models with their background theories and the experimental data is, what it means for a model to be empirically inadequate and so on.

Now within a chosen linguistic framework for models, say an abstract-objects-language, further questions arise regarding the *existence* and the exact *nature* of models qua abstract objects. These questions are internal to the framework and thus they are theoretical. Given that one has accepted an abstract-objects-language for models, the question 'Do these abstract models exist?' is trivial and the answer is of course positive.[8] The further question of the exact nature of these models depends on the internal rules of the framework and the introduced mechanisms for ascribing physical properties to abstract entities.

As an example, consider once again the familiar case of the ideal pendulum in classical mechanics. Introducing a framework which sees the ideal pendulum as an abstract object implies that the further (internal) statement 'there exists an abstract object which has all the properties of the ideal pendulum' is trivially true in the sense that such an object is an element of the chosen framework. However, claiming that the abstract ideal pendulum exists, does not amount to any sort of ontological commitment of a Platonic nature simply because it is not an external statement regarding the *real* existence of such entities independently of the chosen framework. Nor does the fact that such a language may be proven extremely efficient provide any sort of evidence towards these claims. Rather, it merely 'makes it advisable' to accept the specific framework in the sense that it provides all the necessary conceptual and linguistic tools to understand certain aspects of scientific modelling, such as the fact that physicists do indeed seem to refer to abstract objects with spatio-temporal properties when talking about ideal pendulums and frictionless inclined planes.

The further reading of [Q] and other related questions as external theoretical questions is not a viable option. For the Carnapian philosopher of science, this reading of [Q]

---

[8] Things become a bit more complicated in the case of fictionalism, since this doctrine explicitly denies the existence of abstract entities. Part of this complication stems from the fictionalist's failure to acknowledge the distinction between internal and external statements, making fictional statements seem contradictory in the sense that although models do not exist they possess physical properties. We shall return to this issue in Section 2.2.3.

as a question of the *real* ontology of scientific models, over and beyond any linguistic framework we may use to describe them, is both misleading and unnecessary. It is misleading because, as a supposedly theoretical question, it implicitly assumes that there is a definite answer to the question of the ontology and other related questions on the existence and the metaphysical nature of models. However, insofar as such questions cannot be formulated in a way that renders them intelligible and for which an efficient methodology can be suggested towards their resolution, they remain pseudo-questions and thus should be discarded. It is also unnecessary because the alternative understanding of these questions as external practical questions or internal questions within a chosen framework is sufficient for a fruitful explanation of the practice of scientific modelling.

From a Carnapian point of view, there is thus nothing more to be gained by pursuing metaphysical questions about the existence and the *real* nature of the abstract entities that are often found in theories of models, and the burden of proof is on those who suggest otherwise. Namely, they need to make clear what the extra benefits of pursuing metaphysical (external theoretical) questions are, compared to the proposed Carnapian reading which remains completely neutral as to any kind of ontological commitments in the traditional metaphysical sense. The main advantage of the suggested Carnapian take on the question of the ontology of models is that it paves the way for making progress in understanding the function and nature of scientific models by answering the question of the ontology in an internal sense. We thus have no compelling reasons to consider the pursuit of further metaphysical questions about scientific models as a worthwhile task. What is at stake in the long-standing debate on the ontology of models is not a conclusive answer to the question *per se*, rather, the extent to which the different choices of language illuminate different aspects of the nature and function of scientific models.[9]

As for the further questions [1] and [2], it should be clear by now that the answer

---

[9] Based on Quine's (1951) famous response to Carnap, one may express a neo-Quinean objection at this point, arguing that internal questions are ultimately just as pragmatic as external ones, and hence, there is no definite internal answer to the question of the ontology of models either. As mentioned in the introduction, Carnap's meta-ontological stance is taken as a working premise and the defence of Carnap's programme against well-known objections like this one is beyond the scope of this thesis. However, let us just briefly note once again that the aim of this chapter is to show that the reading of [Q] as an external theoretical question is partly responsible for the lack of progress in the debate about the ontology of models. This conclusion is based on Carnap's doctrine that external questions are ultimately practical questions, which essentially remains unharmed by Quine's claim that internal questions are also pragmatic. The further claim that there is no definite internal answer to the question of the ontology of models even within a chosen linguistic framework is also orthogonal to the argument provided here. Whether or not an internal ontological claim is ultimately a pragmatic issue does not really affect the main claim of this chapter, namely, that the question of the ontology of models should only be taken as a means of probing other related questions on the function and nature of scientific models as epistemic tools.

to these questions depends on how one reads [Q]. Recall that question [1] asks whether we can provide an answer to the question of the ontology of models [Q] and question [2] whether answering this question is necessary for our purposes as philosophers of science. If [Q] is seen as an external practical question, then [1] merely asks whether we can come up with an appropriate linguistic framework that captures scientific modelling and the answer is of course positive. Similarly, the answer to [2] is also positive since if our aim is to understand what models are, we of course need an efficient linguistic framework to describe them. If [Q] is seen as an internal question, then [1] asks whether we can describe the nature of models *within* a particular linguistic framework, and the answer to this question is again positive and follows from the specified internal rules of the preferred framework. For the same reasons, the answer to question [2] is trivially positive as well.

If [Q] is seen as an external theoretical question, then question [1] asks whether we are able to determine the ontology of models in a language-independent way and the answer is negative since, from the Carnapian point of view, it is simply impossible to provide an answer to an external theoretical answer. Similarly, question [2] asks whether it is necessary to answer these external theoretical questions about models in order to understand and explain the general practice of scientific modelling and the answer is again negative, since the main motivation of the Carnapian approach is precisely the claim that external theoretical questions do not pose any serious concerns towards our philosophical understanding of various issues. Rather, they often have the opposite effect of impeding our philosophical enquiries. The upshot is that the answer to the two meta-questions [1] and [2] arising from French's discussion of the ontology of models depends on how one understands the central question [Q]. Reading [Q] as an internal theoretical question or an external practical question allows a positive answer to [1] and [2], whereas reading [Q] as an external theoretical question makes the answers to these questions negative.

This does not amount to an outright quietism about the ontology of models however. It is simply a reminder that ontological questions about scientific models do not lie within the sphere of metaphysics. Models are not 'creatures of darkness', as Quine (1956, p.180) once called 'intensions' and other non-physical entities, and the question of their ontology is not a metaphysical matter. Rather, they are epistemological tools used by scientists and the answer to the question of what models are is to be found in the domains that they are being practically used and studied by scientists, that is, in textbooks, labs, conferences, scientific papers and even verbal discourse. This Carnapian approach on the ontology of models takes references to abstract entities as mere linguistic tools – i.e.

internal assertions – within a given linguistic framework, and refuses to engage with any sort of metaphysical enquiry regarding the existence or non-existence of the abstract entities in question. This might give the impression that the suggested view shares much in common with other anti-metaphysical approaches in philosophy such as fictionalism and agnosticism, but this is not the case. The next section highlights the differences between these views.

2.2.3 Alternative ways of rejecting metaphysics

A common strategy to avoid metaphysical enquiries into the nature and existence of abstract entities is to simply deny their existence and regard them as fictions. Consider the question of the ontological nature of models. A fictionalist will typically hold that talking about models is nothing but a useful fiction, and saying that a model – e.g. the ideal pendulum – has such and such properties is, strictly speaking, a falsehood.[10] There is no such thing as 'the real nature' of models, rather there is only the nature of a model 'within a game of make-believe' which is a matter of pretence and has nothing to do with how the world actually is. Hence, since abstract entities do not exist, there is no need for any kind of metaphysical enquiry for the existence or the real nature of models, and thus fictionalism can be seen as a possible way of rejecting metaphysics.

The idea of models having a nature within a game of make believe seems prima facie very close to the idea of models having a nature within a linguistic framework. This is true, however, notice that fictionalism is a strictly antirealist view on metaphysics since it involves a strong metaphysical claim, namely, the denial of the existence of abstract entities. Rather than remaining neutral on the external question of the existence of abstract entities, the fictionalist makes an assertion by claiming that the external statement that there are no abstract entities is true. This is not the case for the Carnapian pragmatist however. Such external statements are simply non-cognitive and thus they should not be seen as truth-apt assertions. The reason is simple. If the fictionalist is pressed to explain what she means by saying that abstract entities do not exist, just like anyone who is a realist about abstract entities, she cannot provide any sort of possible evidence or justification for her claim which will make it seem non-arbitrary.

---

[10] A word of caution is in order here. The fictionalist described here is one who holds an antirealist view on the existence of abstract entities, such as Frigg (2010), Frigg and Nguyen (2016) and Toon (2012). However, although fictionalism and antirealism about abstract entities usually come together, not all fictionalists on models are antirealists about their existence. For example, Godfrey-Smith (2006) although sympathetic to fictionalism on models, seems to hold an agnostic stance with respect to their existence as abstract entities.

What are fictions then for the Carnapian pragmatist? Carnap only mentions fictionalism in passing, saying that fictional statements are false internal statements (2012, p.254). What this means, is that fictional statements are not part of the set of statements that one accepts as true when one adopts a linguistic framework. For instance, within the numbers-language, an even number which is both greater than three and prime is a fictional number, since according to the rules of this framework it does not exist (in an internal sense). In the context of scientific models, fictionalism can be seen as a theory which accepts a linguistic framework whose rules do not entail the (internal) existence of abstract entities. Nevertheless, the important thing to keep here is that the essential difference between fictionalism and the suggested pragmatic approach is that the former includes external assertions whereas the latter does not.

Another possible way of bracketing metaphysical questions is to take an agnostic stance. That is, one may commit to the thesis that metaphysical existential claims about the real ontological nature of models do indeed have a definite objective truth-value, but this truth-value is, nonetheless, not ‘false’ as the fictionalist says, rather it is in principle unknowable. Hence, the best we can do is to *accept* (in a sense that does not imply to believe as true) one claim over another based on certain criteria such as simplicity, ontological economy, accordance with intuition and explanatory power. Which criterion is the most important usually depends both on the specific purposes of an enquiry and the personal standards of the enquirer, and thus, this approach carries an element of subjectivity. Whereas a philosopher might consider simplicity as the ultimate desideratum of a metaphysical theory, another might embrace a theory which, although ontologically inflationary, has greater explanatory force. Nevertheless, as opposed to the antirealist, the agnostic does not make any ontological assertions, but instead chooses to remain neutral as to the definite answer of the metaphysical issues in question.

This view shares more in common with the suggested pragmatic approach, in the sense that it remains neutral to the answer of the question of the external existence of abstract entities. There is, however, a fundamental difference. Although the agnostic refuses to make a metaphysical claim about the existence or non-existence of abstract entities, she takes external statements to be truth-apt assertions for which we merely do not know whether they are true or false. Again, this is at odds with the Carnapian approach, since for the Carnapian, these questions are neither true nor false.

In all, the Carnapian view, is not an antirealist view that denies the existence of abstract entities, nor an agnostic view that takes metaphysical statements as truth-apt assertions whose truth-value is in principle unknowable. Rather, it is a third more sophisticated view which rejects both the realist - antirealist dichotomy on metaphysics and the need to ask metaphysical questions at the first place. This nihilistic view on metaphysics can be either global and thus apply to any kind of metaphysical enquiry, or local, based on the observation that a certain metaphysical question presupposes a misleading and false usage of language for a particular domain of inquiry. And as already noted above, the use of terms that seem to refer to abstract entities is merely an act of compliance to the use of an appropriate linguistic framework, in lack of any alternative and equally efficient way to describe and explain a given issue with no reference to abstract entities.

Macarthur and Price (2007, p.99) have nicely summarised this pragmatic 'no-metaphysics' view as follows:

> Our pragmatists are [...] happy to stand with the folk, and to affirm the first-order truths of the domains in question – to affirm that there are beliefs, and values, and causes, and ways things might have been, and so on. What they reject is any distinctively metaphysical theoretical perspective from which to say more about these matters – that they do or don't *really* exist, that they are *really* something subjective, or whatever.

Going back to the ontology of models, we the Carnapian pragmatists are happy to affirm that models are abstract incomplete objects or mathematical structures, or fictions, however, in the absence of an appropriate theoretical or scientific perspective, we refuse to engage with the question of whether these things *really* exist or not, or whether there is a single true answer to the ontology of models. The next section explores French's quietism and shows how this approach lies within the sphere of the suggested pragmatic anti-metaphysical approach. In accordance with pragmatism, French denies that there is a unique *true* answer to matters of the ontology of models and suggests that the way forward is to choose the 'ontology' which best represents scientific models without worrying if this ontology is actually true.

2.3 French's quietism

As already noted in Section 2.1, French's quietism stems from two major claims. The first claim is that questions about the ontological status of models are unanswerable, in the sense that no unique and true answer can be given which covers all kinds of models. In other words, for French, questions regarding the real ontology of models *cannot* be answered. The second claim takes a step further and asserts that the inability to arrive at definite answers to these questions should not concern us since it does not impede our efforts as philosophers of science. That is, we *need not* answer these questions in order to understand the function and nature of models. The conclusion is therefore that, rather than searching for an objectively true answer, one should focus instead on finding the most appropriate way to *represent* models and theories as having a certain ontological status, based exclusively on pragmatic grounds. In what follows, French's main argument towards a pragmatic view is evaluated and found susceptible to a number of objections. The upshot is that a pragmatic approach cannot and need not be based on the fact that the term 'model' is not a sortal term.[11]

2.3.1 Models and sortals

French's main argument towards our inability to answer ontological questions about scientific models is based on the concept of *sortals*. Its structure can be given as follows:

[A1] The terms 'theory' and 'models' are not sortal terms.

[A2] Ontological questions about terms that are not sortal are unanswerable.

[A3] Therefore, ontological questions about theories and models are unanswerable.

In general, a term is sortal only if it gives a criterion of identity and countability about a thing. That is, if X is a sortal term, then when confronted with instances of X, one should be able to both identify them as Xs and count them. For example, the term 'owl' is a sortal term since it is clear which entities count as owls and which not, whereas terms such as 'gold' or 'heap' are not, since the former is uncountable and the latter has no clear identity conditions.[12] Moreover, according to some views, a sortal also tells us when something continues to exist and when it goes out of existence. Sortals are therefore terms that designate entities for which identity and persistence conditions are

---

[11] French's discussion equally revolves around both theories and models since the two terms are mostly used interchangeably throughout the text. The main reason for this is French's belief that the function and nature of models and theories cannot be sharply distinguished (*ibid.*, p.241). In what follows, the discussion is limited to models, assuming, rather safely, that even though it is not made explicit in the text, most of French's claims about theories apply to models as well, and vice versa.

[12] The terms countable and uncountable are used here in the ordinary grammatical sense and should not be confused with uncountability in set theory. Countable nouns refer to discrete objects that can be counted – e.g. owls, electrons, planets etc. – whereas uncountable nouns stand for things that are treated as an undifferentiated unit, rather than as something with discrete elements – e.g. gold, electricity, music etc.

clearly determined. Consequently, sortals typically refer to entities of a single ontological kind and therefore, ontological questions about sortals are easier to pursue.[13]

French's starting point for justifying [A1] is the observation that when we ask questions like 'What is the ontological status of theories and models?' we are treating these terms as sortals, since what we are doing is to '[take] the term theory [or model] and ask what it is that this term picks out, what is its referent' (*ibid.*, p.240). However, the great heterogeneity of different types of models makes it impossible to define what a model is in terms of necessary and sufficient conditions, and therefore we lack the desired identity criterion. Just like the term 'works of art', for example, covers too broad a spectrum of an entity (e.g. novels, paintings, sculptures etc.) so does the term 'scientific models', and thus the question of the ontological status of models is unanswerable in the sense that there is no unique answer (*ibid.*, p.241). Moreover, French points out that whether or not one aims for a unified answer to the question of the ontology of models depends on one's understanding of models and their relationship with theories. An understanding of models as some sort of extensions of theories suggests for a single and unified answer, whereas an understanding of models as having a different nature and function than theories, such as in Cartwright *et al.* (1995) and Morrison and Morgan (1999), suggests that models and theories refer to two different things for which different answers should be given (*ibid.*, p.242).[14]

The same can be said for the persistence criterion, since it is not clear when a theory comes into existence and when (and if) it ceases to exist. French wonders: 'did General Relativity just pop into existence when Einstein thought it up? And when exactly did he do that? Did it partially come into existence in October 1914 and only fully the next year after Einstein's correspondence with Hilbert?' (*ibid.*, p.239). Replace General Relativity with Bohr's model of the hydrogen atom and the same argument holds for models. Did Bohr's model come into existence partially as he was gradually developing it? Or did it suddenly come into existence with the publication of his paper in 1913?

[A2] is supported by the fact that the scientific practices which are supposed to determine the identity and persistence conditions of models draw no sharp lines on

---

[13] There are various views in the literature as to where the term 'sortal' applies (universals, concepts or the things themselves) and French is not explicit on which interpretation he adopts. Following Quine (1960), the term 'sortal' will be treated here as a linguistic notion applying to predicates, since this approach is compatible with French's overall discussion.

[14] Interestingly, one might say here that the proposed account in Chapter 1 combines the best of two worlds, in which models are indeed extensions of the theory in that they are specific applications of more general theoretical principles, but at the same time they perform a different role and function as epistemological tools providing insights for the physical phenomena they represent.

whether something should be seen as a model or not. For instance, they do not tell us how much of a model could be altered in order for it to remain the same model, or when the model comes into existence. Hence, this lack of any determinate conditions of identity and persistence makes it hard to see how we can arrive at some determinate answers. In other words, one cannot say what the ontological status of a model is, if one is not sure what the referent of that term is or when the term actually refers to something.[15]

There are two possible ways of response to this argument by challenging each one of its premises. First, one might reject the concept of sortals as an ill-defined concept and press for a definite answer to the question of what exactly makes a predicate a sortal term. Is it the fact that there are clear identity criteria for the term's referents or the ability to distinguish certain things as being instances of that predicate? One might argue for example that even though no clear identity criterion or criteria for what counts as a model can be formulated in terms of necessary and sufficient conditions, surely it is still possible to distinguish and count different cases of models. For instance, physicists have no problems in distinguishing the Fermi gas model of the nucleus from the shell model; and as a matter of fact, there are over thirty different models of the nucleus, each based on different assumptions, which can nonetheless be classified in various ways.[16] It is therefore possible, at least in principle, to identify and enumerate all cases of models in physics say, or even all cases of models across all scientific fields, by making a long open-ended list and leaving any ambiguous cases aside. Once this list is done, one may take its contents as the referents of the term 'models' and thus treat the term as a sortal.

What is more, the desirable identity and persistence conditions given by sortals turn out to be problematic even in cases which prima facie seem clear examples of sortals, such as the term 'apple'. This is because, just as in the case of models and theories, the spatial and temporal boundaries for something to be considered as an apple are not as clear as one might first think. To see why, compare French's questions on the identity and persistence conditions of General Relativity with questions on the identity and persistence of apples. When does an apple come into existence? Does it come partially as it develops from a blossom into a hard mass fruit? If no, at what time then does it stop being a blossom and count as an apple? And how big of a bite can someone take

---

[15] What is presented here is a summary of French's argument as it appears throughout Section 3 of his paper (pp.238-243), which relies heavily on Thomasson's (2006) discussion of the ontology of art.

[16] See Greiner and Maruhn (1996) for a book-length classification of nuclear models based on degrees of freedom.

after which the apple stops to exist?[17] If even in these simple cases no clear identity and persistence conditions can be given, it is then hard to see when a term successfully counts as a sortal, and more importantly, it is even harder to see why it is a necessary condition for a term to be sortal in order to ask ontological questions about its referents as [A2] implies.

Even if we accept a certain definition for sortals, and grant that models and theories are not sortal terms, we can thus still question the second premise of the argument which after all carries the most important weight. That is, we can deny that it is a necessary condition for a term to be sortal in order to ask ontological questions about its referents and thus deny that ontological questions about models are unanswerable. Take the term 'gold', for instance. Even though it is not a sortal term according to the above definition, it is clear that one can still answer ontological questions regarding the nature of gold. What is more, even if we accept that non-sortal terms such as 'works of art' and 'models' refer to entities of various ontological kinds, one might maintain that different classes of models pick out objects of different ontological kinds, but nonetheless we can categorise these kinds and make separate ontological claims for each one of them. This is the line followed by Contessa (2010) for example, who argues that models should be categorised in three kinds – material models, mathematical models and fictional models – for which questions about their ontological status can be answered separately.

French is fully aware of this possibility, hence his conclusion is not that the question of the ontological status of models is inherently unanswerable *tout court*, rather it is the much weaker claim that it is unanswerable in the sense that no single unified answer of the form 'all models are F's', where F is a specific ontological kind, can be given. It is hard to see how this leads to quietism however. The fact that several answers can be given to the question of ontology does not imply that the question cannot be answered. French's observation that models are not sortal terms nicely demonstrates the vast array of scientific models and the unsystematic use of the term by scientists, which make the task of developing a comprehensive theory of models extremely difficult. However, as an argument towards quietism it suffers both from the fact that the concept of sortals is ill-defined and from Contessa's alternative tripartite approach. The argument thereby does not succeed in showing that the question of the ontology of models is unanswerable, nor does it show that it is not worth pursuing. The good news however, is that all French needs in order to defend the stronger claim that questions about the real ontology of

---

[17] This argument against the temporal and spatial boundaries of the extensions of predicates is found in Teller (2018), although in a completely different context, in a discussion of the inaccuracy of human knowledge.

models cannot be answered, is the Carnapian rejection of the disguised external questions as pseudo-questions.

Once this is done, all we need for quietism to follow is to show that answering these questions is unnecessary. French easily achieves this by developing an argument based on the work of Peirce (1940) showing that external ontological questions about the *real* nature of scientific models are not genuine questions since they do not impede in any way our enquiries as philosophers of science (*ibid.*, pp.243-4). The upshot is that a fruitful theory of scientific representation does not require any kind of metaphysical assertions about the existence of abstract entities. What is needed is a moderate representational attitude guided only by pragmatic criteria. Whether one finally concludes that models are best seen as mathematical structures or fictional objects, or as consisting from the four main features presented in the previous chapter, is merely a result of a pragmatic choice based on the ability of the competing theories to explain the nature and function of scientific models in the best possible way, admitting as few counterexamples as possible.[18]

Following these observations, French's quietism does not seem to be as radical as one might first think. Instead, it can be interpreted as stating that external theoretical questions about the real ontology of models do not hamper our efforts towards developing a theory of models since they can be replaced by external practical questions and internal questions within a chosen linguistic framework. Once this premise is granted, quietism about the metaphysics of models follows naturally.

2.4 Thomson-Jones against the bracketing of metaphysics

In this last section a possible objection to the proposed view on the ontology of models is addressed. This objection comes from Thomson-Jones (2017, pp.244-5) who, as opposed to French, argues extensively that bracketing metaphysical questions in philosophy of science impedes our overall understanding on issues like the ontology of models and scientific representation. In order to fully appreciate his argument, consider a theory [T] containing a statement [t] referring to abstract objects which, nonetheless, remains neutral as to the existence of these objects:

[t] Scientific models are abstract objects.

---

[18] French (2021) reinforces this view in a more recent paper.

By formulating theories in this way, one is engaging with what Thomson-Jones calls the 'as-if practice', namely the practice of talking as if there are X's (in this case abstract objects) and as if they have certain features (*ibid.*, p.234). Thomson-Jones's argument then proceeds as follows:

[B1] Either there are abstract objects such as the simple pendulum or not.

[B2] If there are, then [t] should be taken literally.

[B3] If there are not, then scientific modelling does not involve such objects and therefore [t] should not be understood literally.

[B4] If there are no abstract objects but [t] is true nonetheless, then it is not obvious what [t] means.

[B5] If we do not know whether there are abstract objects, we cannot know whether the account of modelling is to be taken literally.

[B6] Therefore, we cannot claim to have arrived at an understanding of modelling by invoking such an account in the midst of such a fundamental uncertainty about the actual meaning of [t].

[B7] Removing that uncertainty will at least involve answering the existence question about abstract objects.

[B8] Therefore, bracketing is not an available option.

As it stands, the argument is supposedly directed against all possible ways of bracketing metaphysics in philosophy – i.e. by taking an agnostic stance towards metaphysical existence related questions, by explicitly denying the existence of abstract entities like the fictionalists do, or by taking a Carnapian approach. The gist of the objection is that no matter which approach one takes for bracketing the (external theoretical) question of ontology, [t] is always left unexplained. This is because [T] is an attempt to explain what models are and how they are related to their targets by involving talk of abstract entities. Therefore, the (external) ontology of these entities plays, according to Thomson-Jones, an important role. This is reflected in [B1] which echoes what Thomson-Jones calls the 'existence question' about abstract objects. To claim that they do exist, is to make an ontological commitment and thus – as [B2] shows – [T] as a theory of models provides an understanding and a possible true explanation given that [t] is true. However, any attempt to refuse engaging with the metaphysical question of the existence of abstract entities leaves us with uncertainty as to the actual meaning of [T] and thus, according to Thomson-Jones, provides little understanding. [B3] and [B4] clearly aim for the fictionalist, and [B5] targets the agnostic approach.[19]

What about the Carnapian approach however? Thomson-Jones does not engage with this option in detail, and the reason is that he presupposes that the (external theoretical) existence question, on which [B1] relies, is a legitimate question to ask. That is, he presupposes that it is a matter of fact that abstract entities either exist or not. However, this is exactly what the Carnapian pragmatist denies and thus the argument breaks down at its very starting point. For the Carnapian pragmatist, the external theoretical question of the existence of abstract entities is a non-cognitive pseudo-question. Insofar as this question cannot be formulated in a way that makes it cognitively intelligible, it is simply inappropriate and it should be discarded.

Thomson-Jones justifies [B1] by saying that 'when evaluating an account which engages the as-if practice for X's [e.g. abstract objects], it is prima facie entirely reasonable to ask, as part of the evaluation, whether there are indeed X's, and if so, whether they are the right sort of thing to play the roles the core account would seem to require them' (*ibid.*, p.248). But this assumption only leaves the Carnapian wondering. What does it mean for an abstract entity to exist? And how can we ever tell whether an abstract entity exists or not? More importantly, what is the difference between an existing abstract entity and a non-existing abstract entity? Until we find an appropriate way to answer these questions in a meaningful and constructive fashion, they cannot be considered as legitimate, let alone as an indispensable part of a theory of scientific models.

In fairness to Thomson-Jones, he clearly states that he is not arguing that our philosophical enquiries should be put on hold until we reach a definite answer to these questions. What he is arguing for is that we have to acknowledge that the answer to one question (say to the question of the ontology of models) 'depends in part on the answer to a number of other, equally difficult and uncertain questions' (*ibid.*, p.234). And a sensible way of coping with such difficult situations is to make a working hypothesis, a sort of 'educated bet', and develop our theories based on that assumption. One is left wondering however, whether there is any practical difference between this educated guess about the nature of models and the introduction of what one takes to be the most efficient linguistic framework for a given aim.

---

[19] It is not my purpose to defend a fictionalist approach to modelling here, however it is worth mentioning that with regard to [B4], this is exactly what the fictionalist's theory aims to explain by appealing to pretence and games of make-believe.

Insofar as external ontological questions cannot be formulated in common scientific language in a way that makes them cognitively intelligible, to introduce a tentative answer to such questions – say to make a working assumption of the sort 'models are existing abstract objects' – looks more like giving a pseudo-answer, as Stein has aptly noted (1989, p.54); and it is highly doubtful whether such claims provide the kind of understanding Thomson-Jones is seeking for, according to his own principles. Recall that Thomson-Jones' criticism to the fictionalist (premise [B4]) is that given that [T] contains a claim which is literally false, an important part of this theory remains unexplained, or even worse, false. Does the introduction of an external assertion as a working hypothesis make things better however? Stein's point is that what is actually happening in these cases, is that a supposedly explanatory notion is introduced which when examined carefully is found to be in effect completely disconnected from the explanandum (hence the 'pseudo-answer'). In other words, given our inability to provide a robust meaning to such metaphysical existential claims, these claims fail in providing a satisfactory explanation as part of our theory. The solution is to see the hypothesis of the existing abstract entity merely as a linguistic tool which facilitates our talk of scientific models, and not as a serious ontological commitment in the metaphysician's sense.

2.5 Conclusions

This chapter is a result of the observation that a significant part of the literature on modelling and scientific representation concerns the metaphysical implications of the debate on the ontology of models. This fact gives the further impression that these matters are closely associated with a number of persisting problems in traditional metaphysics, such as the existence of abstract objects and the nature of properties. Following Carnap, the suggestion here is to see the question of the ontology of models as either an internal theoretical question within an already accepted linguistic framework or an external practical question regarding the choice of the most efficient theory in order to explain and understand certain features of scientific models. The choice between competing theories of models therefore depends solely on the relevant pragmatic criteria and the specific desiderata of each account.

The four-features account presented in the first chapter of the thesis is precisely a useful linguistic framework for understanding the role of models in science as epistemological tools. The desiderata for presenting such an account stem from the overall theme of the thesis as a model-based philosophical investigation of the methodology of modern physics and a critical analysis of the relationship between models and theories, the relationship between models and experimental data, and the role of models in the observation of unobservable physical entities such as dark matter particles. The aim of the present chapter was to show how once one adopts a pragmatic stance, one can dispense with various metaphysical conundrums that otherwise seem insurmountable and employ a useful linguistic framework for pursuing further questions about the role of models in the methodology of science.

The main implication of this Carnapian view is therefore that the question of the ontology of models is merely a means of probing other related questions regarding the overall practice of scientific modelling and the function of models as epistemological tools for gaining knowledge about the physical world. One of these questions, which will be the focus of the next chapter, concerns the relationship between theories and models and the ways in which the latter can be derived from the former.

Finally, let us also note that the framing of the debate on the ontology of models in Carnapian terms nicely illustrates how Carnap's approach is still relevant for contemporary discussions in the philosophy of science and that several lessons can be drawn from it. Perhaps the most important lesson to be learned is that before setting out to answer a philosophical question, we should first pause and think what the question is really asking and what we seek to understand by exploring the possible answers to it. This way we can avoid 'the danger of getting into useless philosophical controversies' (Carnap, 1935, p.76) that Carnap was trying so hard to abolish. A fruitful debate is one in which all parts have a clear and common understanding of the problem in hand, and the nature of the debate on the ontology of models shows that this might not be the case. What is being put forward here by appealing to a Carnapian take on the debate is not an outright quietism about the ontology of models. Rather, it is a gentle reminder that there is nothing to be gained by trying to settle down to a unique answer on the question of the ontology of models.

CHAPTER 3

THEORIES WITHOUT MODELS

PERTURBATIVE QUANTUM FIELD THEORY MEETS CONSTRUCTIVE EMPIRICISM

In our attempt to distinguish models from theories in the first chapter, we defined the latter as the primary and comprehensive frameworks that provide the fundamental principles capturing the behaviour of a class of different yet similar phenomena. We have also seen that a well-defined physical theory, such as quantum electrodynamics, can be expressed by a set of equations, laws and fundamental principles that make its *hard core*, and a set of auxiliary claims. And for each one of these theories, there is a set of models which can be derived from the theory alone, without taking into consideration any kind of empirical data. Following van Fraassen we labelled these as the *theoretical models*. For van Fraassen, together with the observable phenomena, theoretical models occupy the central stage of constructive empiricism, since these are 'the two poles of scientific understanding, for the empiricist [..] The former are the target of scientific representation and the latter its vehicle' (2008, p.238).

The main goal of this chapter is to argue that the importance of theoretical models in physics, as described by van Fraassen, is significantly undermined by the modelling practice of perturbative Quantum Field Theory (QFT) which relies heavily on regularization and renormalization techniques.[1] The central claim is that the models produced via these techniques in high energy physics experiments are not theoretical, but nonetheless remain an important and indispensable tool for making progress in this field. The

---

[1] For philosophical discussions on renormalization see Batterman (2001), Butterfield (2014), Reutlinger (2014), Reautlinger and Saatsi (2018), Rivat (2019) and Williams (2015).

outline of the argument in support of this claim is the following: Theoretical models are models that are deductively derivable from theories, or models that are products of controllable idealizations. Models of perturbative quantum field theory are neither deductively derivable from theoretical principles nor are the results of controllable idealizations. Therefore, the models of perturbative QFT are non-theoretical.

The concept of 'controllable idealizations' is borrowed from Sklar (2000, pp.44-5) and concerns the fact that, during the construction of models, physicists often resort to various mathematical techniques and approximations for which no rigorous mathematical or physical justification can be given. An idealization is controllable if the scientist has all the necessary resources available by the current scientific knowledge to explain why the effect of the introduced assumptions in a model is – up to a certain degree – negligible. Ideally, the scientist is also in a position to provide estimates of the deviation between (i) the experimental results, (ii) the theoretical predictions of the idealized model and (iii) the exact solutions of the original mathematically intractable theoretical model. For our purposes, the lack of a rigorous mathematical or physical justification for the introduced assumptions during the construction of a model is understood as an act of introducing uncontrollable idealizations.[2]

The first premise of the argument is a necessary assumption in order to make clear the distinction between theoretical and non-theoretical models. Generally speaking, theoretical models are deductively derivable mathematical structures in which the dynamical equations and the fundamental laws of a physical theory are completely satisfied. However, as we shall see, theories often produce models with intractable mathematical equations, and hence further mathematical simplifications are necessary in order to derive the numerical predictions of the model. If the assumptions introduced during this process are controllable, then the resulting model is still a theoretical model, although – strictly speaking – it is not deductively derivable from the theory. On the contrary, models in which the dynamical equations and fundamental principles of the background theory are not satisfied are considered to be non-theoretical.

The study of the central aspects of the modelling practice in perturbative QFT shows that some of the introduced idealizations in highly successful models in this field are not controllable, due to the fact that the involved approximation methods are often

---

[2] Sklar's concept of controllable idealizations can also be linked to what is often discussed in the literature as Earman's Principle: 'While idealizations are useful and, perhaps, even essential to progress in physics, a sound principle of interpretation would seem to be that no effect can be counted as a genuine physical effect if it disappears when the idealizations are removed.' (Earman, 2004, p. 191). For further discussions of this principle see Landsman (2013), Ruetsche (2011, Ch.14) and Fletcher (2020).

mathematically unjustified. In what follows, it will be shown that certain steps during the construction of empirically relevant models based on perturbative calculations, cannot be justified by offering an accompanying explanation for why the resulting product is an appropriate approximation of the problem in question. This fact automatically breaks the strong deductive relationship between theories and models required by the semantic view of theories as it is implemented in van Fraassen's constructive empiricism.

As we have seen in the first chapter, the autonomy of models from theories has been discussed by several authors in the past, mainly as a criticism to the semantic view. For instance, Portides (2005) raises an important objection against the semantic view of theories using the liquid drop model of the nucleus as an example. By showing how the model was introduced in the 1930s to explain – amongst other things – the Weizsäcker semi-empirical formula of the binding energy of the nucleus, Portides argues that certain crucial assumptions of the model lack the necessary theoretical justifications which would allow one to consider the liquid drop model as a theoretical model (*ibid.*, p.1294). In a similar spirit, Cartwright, Suarez and Shomar (1995) and Morrison (1999) argue for the autonomy of models from theories using the examples of the Londons' model of superconductivity and Prandtl's model of ideal fluids respectively.

What is common in these accounts is the appeal to examples of phenomenological models in physics. The main characteristic of phenomenological models is that their construction is primarily guided by empirical observation and experimental data, with minimal dependence on the available theoretical framework. As a result, phenomenological models often violate the basic theoretical principles and fundamental laws of the background physical theory, and hence, the claim is that such models are often independent and autonomous from background theoretical principles. While this is true, the claim I wish to make here is that there is an additional type of models in physics – for which the models of perturbative QFT are a prime example – in which the strong deductive relationship between theories and models breaks down. The reason however, is not the fact that these models are primarily guided by phenomenological considerations. Rather, it is the fact that the mathematical treatment of these models results to the introduction of uncontrollable idealizations due to the lack of a rigorous mathematical and/or physical justification.

This third type of models bears a strong resemblance with Redhead's (1980) *floating models* discussed in Section 1.3.3. For Redhead, floating models are models that are disconnected from a background theory due to a *computational gap* arising from our inability to justify the validity of the mathematical approximations used for the construction of the mathematical framework of the model. These models also fail to provide successful predictions and thus, they are, in a sense, ‘floating’, since they are detached both from theory and the experimental data. Nevertheless, the value of floating models is that they serve as the preliminary steps for probing various essential features of the theory and exploring the possible ways of refining the theory in order to yield theoretical models that are predictively successful. However, the main difference between Redhead’s floating models and the examples of perturbative QFT to be discussed here is that the latter exhibit a remarkable agreement with experimental data despite the fact that – as products of uncontrollable idealisations – they are detached from the theory by a computational gap. This suggests that the distinction between theoretical and phenomenological models does not capture all types of models in physics. In addition to what one may call, floating models, there is a third type of models that are neither theoretical nor phenomenological, but nonetheless play a crucial role in modern physics. The models of perturbative QFT are a prime example.

This lack of mathematical rigour in the methods of perturbative QFT has recently attracted the attention of many authors, mainly with respect to the possible interpretations of QFT and its status as a fundamental theory. Based on a notorious result by Haag, Hall and Wightman (known as Haag’s theorem), Earman and Fraser (2006) argued that the interacting models of QFT rest on an inconsistent set of assumptions. As we will shall see, this plays a major role in the relationship between the final empirical models and the background theory, since the various processes for constructing empirically relevant models in QFT often violate the basic principles of the theory that are needed to prove Haag’s theorem, a point made salient by Miller (2018). Doreen Fraser (2009) has further argued that the process of taking cut-off limits in these models is ill-defined, making the interpretative analysis of the theory on a fundamental level difficult, if not impossible. In a similar tone, Wallace (2006) claims that QFT is an ‘intrinsically approximate’ theory of some unknown deeper theory, which gives a mathematically self-contained description of the high-energy/short-distance physics, and as such, it makes no claims about the way the world is at these scales.

In a more recent work, James Fraser (2020) has suggested that the perturbative method of QFT does not produce any models at all; rather, it is a method for producing approximations of physical quantities like scattering cross-sections. Hence, according to him, given the reliance of the constructive empiricists’ approach on the existence of empirically adequate models, the success of perturbative QFT poses serious challenges for this view. My main aim is to examine the mathematically questionable methods of perturbative QFT with respect to the relationship between models and theories, and argue that these methods do produce models, albeit ones that are not deductively related to the background theory. Hence, the real problem for the constructive empiricist is not the fact that perturbative QFT does not produce any models at all, rather it is the fact that it produces models that are non-theoretical.

The structure of this chapter is as follows. In Section 3.1, the concept of theoretical models as defined by van Fraassen (2008) will be presented in more detail, followed by some remarks on the role of idealizations in the construction of these models. Section 3.2 follows with a relatively self-contained analysis of the modelling practice in perturbative QFT in terms of regularization and renormalization. This will facilitate the discussion to follow in Section 3.3 in which it will be argued that the introduction of cut-offs in the models of perturbative QFT by regularization and their further treatment via renormalization breaks down the deductive relationship between these models and the background theory. Two further complications arising from these processes will also be discussed in less detail. Finally, in Section 3.4 the main implications to constructive empiricism and the semantic view from the existence of non-theoretical models will be evaluated.

3.1 Theoretical models and idealizations

Van Fraassen's emphasis on the theoretical models of a theory is a consequence of his strong views in favour of the semantic view, according to which the structure of a theory is nothing but its class of mathematical models. As already mentioned in Chapter 1, for van Fraassen (1980; 1989), the presentation of a scientific theory consists of a description of a class of state-space types, where state spaces are understood as a generic collection of mathematical objects (such as vectors, functions and numbers) denoting the physical state of a system. Theoretical models are understood in this context as mathematical structures that are deductively derivable from the basic principles of the theory, denoting sequences of states that form a trajectory in the state space over time.

The connection of these models to the physical world is described by van Fraassen (2008, p.168-170) in terms of an isomorphic relationship between the theoretical models and, what he calls, the *surface models* of an experiment, i.e. the idealized models of the raw data obtained from experiments. Unlike surface models, the theoretical models of a theory are deductively derivable from the fundamental principles and laws of a theory and are independent of experimental measurements. In van Fraassen's words '[t]he sense in which a theory offers or presents us with a family of models – the theoretical models – is just the sense in which a set of equations presents us with the set of its own solutions' (*ibid.* p.311). And just like a mathematical equation, a theory may have models that we have not yet discovered, but nonetheless, they exist.

A possible way to understand theoretical models, according to van Fraassen, is to think of them as specifying:

(i) a family N of observables (representing physical quantities) each with a range of possible values

(ii) a set S of possible states (i.e. the model's state space)

(iii) a stochastic response function P_s^n for each n in N and s in S, which is a probability measure on the range of n.

The members of the set N of observables are functions from the model's state space to the real line, assigning values to various physical quantities that are typically represented by the parameters and variables of the model. The state space S comprises the set of all possible configurations of the system's variables, and can be thought of geometrically as an N-dimensional space whose dimensions are defined by the number N of observables in the model. A given state s is thus represented by a point in the state space whose coordinates stand for a possible combination of the values of the physical quantities of the system represented in the model. Finally, the stochastic response function P_s^n is the model's specification that a (direct or indirect) measurement of a given observable n will yield a specific value n_i, when the state is s_i.

The overall idea is that a theoretical model *fits* (or is isomorphically embeddable to) a surface model only if the theoretical model has some state s such that the function P_s^n contains the surface state P of the surface model, relative to the given identification of the measurement setups as measurement of the physical magnitudes n (*ibid.* p.169). A physical phenomenon is thus represented both 'from below' via the surface model and 'from above' via the theoretical model. The fitting of these two models establishes the link of the theory with the natural world and eventually its empirical adequacy.

In Chapter 1 we presented the classical harmonic oscillator as a highly abstract theoretical model that serves as the basis for the ideal pendulum. The dynamic behaviour of the ideal system described in the harmonic oscillator is completely determined by the equations of the background theory and is in full accordance with the theory's fundamental principles. However, purely theoretical models such as the harmonic oscillator are, in general, mathematical artefacts with no tangible connections to the physical phenomena that the theory aims to describe, in that the form of these models is such that it does not allow the direct comparison of the model with the corresponding models of data derived from experiments. Hence, the additional construction of *empirically relevant models*, such as the ideal pendulum, that are based on these abstract theoretical models is often necessary, in order to compare the quantities predicted by the models with the outcomes of experimental measurements. Theoretical models thus provide the mathematical basis on which further empirically relevant models are built with respect to a corresponding target system and/or an experiment. The aim is to obtain theoretical predictions for experimental observables via these new models, in order to test the empirical adequacy of the models and consequently of the background theory.

The process of constructing an empirically relevant model from a theoretical model involves the introduction of two different types of assumptions. The first type of assumptions – call them *Type 1 assumptions* – concerns those cases where the theory provides theoretical models for which no analytic solutions can be found and thus further modifications are required in order to make the models mathematically tractable. An example of a Type 1 assumption is the small angle approximation in the ideal pendulum discussed in Chapter 1 (Sec.1.3.1) in which it is assumed that the pendulum only swings in very small angles θ for which we know that

$$\sin\theta \approx \theta \tag{3.1}$$

The small angle approximation yields a new linear equation of motion which has a well known exact solution and thus allows the mathematical tractability of the model. The crucial point is that insofar as we have good (mathematical) reasons to believe that (3.1) holds for small angles, the introduced idealization is mathematically justified, which means it is controllable. One can therefore still think of the linearised version of the ideal pendulum as a theoretical model, in the sense that it is derivable from the theory modulo some controllable modifications.

The second set of assumptions – call them *Type 2 assumptions* — concerns the relationship between the empirically relevant model and the target system it represents. Given that physical systems are in general too complicated to be accurately captured, and given that a model only aims in describing a specific aspect of the system at a particular scale, a number of assumptions are required to bring the model system and the physical system closer, in order to establish the agreement of theoretical predictions and experimental results. This is a two-way process. On the one hand, the physical system which serves as a target of the model is *misrepresented* in order to facilitate the mathematical tractability of the model. This part typically concerns the choice of the appropriate degrees of freedom to describe the system at a particular scale and the further assignment of inaccurate values to some of the physical quantities of the system represented in the model. On the other hand, the relevant experiments are designed in order to avoid as much noise as possible and the physical system in question is prepared in such a way so that it is as close as possible to the system described by the corresponding model.

Using the homely example of the ideal pendulum once again, the first part of this process corresponds, for instance, to the assumption of a homogenous gravitational field, whereas the second part corresponds to the placement of the physical pendulum into a vacuum in order to minimize any effects from air resistance. The justification for this assumption comes from showing that the variance in the value of the gravitational field with respect to the experimental setup is negligible. The closer the real value of the field is to the fixed value in the model, the better the approximation. Moreover, the scientist is also in a position to 'de-idealize' the model by taking into account a number of additional corrections in order to estimate the deviation between the surface model's results and the theoretical model's predictions making the idealizations controllable.[3] The upshot is that insofar as our current scientific knowledge provides the means for shifting between purely theoretical models and their (idealised or de-idealised) counterparts, the latter still depend on the theory and thus they can also be seen as 'models of the theory'. This is exactly what van Fraassen has in mind when he says that 'if the ideal pendulum is a Newtonian model, and it is de-idealized, then the result is still a Newtonian model. For if it is not, then it is a counter-example to Newton's theory, and hence not a de-idealization' (2008, p.310).

Although quite different in nature, both Type 1 and Type 2 assumptions are often concealed in the literature under the broader notion of idealization.[4] What is important however, is whether the introduced idealizations in the empirically relevant models are after all *controllable*; that is, whether the available scientific knowledge is able to provide

---

[3] See Nelson and Olsson (1986, pp.114-117) for a detailed analysis of the various corrections that can be added on the theoretical model of the pendulum.

[4] McMullin's (1985) seminal paper on idealization touches upon this distinction by distinguishing between 'mathematical idealization' and 'construct idealization'. The discussions in Pincock (2007) and Batterman (2009) mainly concern Type 1 assumptions in which the focus is on the mathematical treatment of the models, while Weisberg (2007), Morrison (2015) and Portides (2021) mainly discuss idealization in terms of Type 2 assumptions, focusing on the conceptual and mental processes of constructing models by distorting or omitting features of the physical target systems.

a convincing mathematical or theoretical justification to explain why the empirically relevant models, despite not being deductive products of the background theory, are still able to provide good approximations for the theoretical quantities in question.

In the following two sections, we shall draw on the modelling techniques of perturbative quantum field theory and argue that in addition to the phenomenological models whose construction aims in accommodating empirical results, there is another important class of non-theoretical models in physics whose construction is primarily guided by the introduction of various Type 1 assumptions with the aim of making the models mathematically tractable. The careful analysis of the modelling methodology of perturbative QFT shows that the empirically relevant cut-off models produced by regularization and renormalization techniques are detached from their corresponding theoretical models, in the sense that some of the introduced assumptions are mathematically and theoretically unjustified. This fact breaks down the deductive relationship between the empirically relevant models and their background theory and thus challenges the delicate picture of the semantic view which sees scientific models as mathematical structures that satisfy the basic principles and equations of the background theory and for which the theory is, strictly speaking, true.

3.2 Models in perturbative Quantum Field Theory

As we have seen in Chapter 1, quantum field theory is the standard theoretical framework for constructing models for scattering processes of subatomic particles in particle physics. However, the current axiomatic formulations of QFT are only capable of providing physically irrelevant models that describe either non-interacting systems or interacting systems in lower spacetime dimensions than the four-dimensional spacetime. Hence, physically relevant models can only be constructed by using scattering theory calculations in Lagrangian quantum field theory, such as the ones presented in Chapter 1 using the specific example of an electron positron annihilation in QED. These calculations allow us to construct empirically relevant models that are capable of providing theoretical predictions for physical quantities that can then be compared directly to experimental observables such as cross sections and decay widths. The first difficulty in this process arises from the fact that the construction of four-dimensional interacting models via scattering theory leads to intractable Hamiltonians since the introduction of interacting terms typically leads to non-linear equations of motion for which no exact solutions can be found analytically. In order to make the interacting models mathematically tractable, physicists often appeal to perturbation theory.

Recall that in Chapter 1, in the context of the specific example in QED, we briefly mentioned that the exact expression of the Feynman amplitudes $\mathcal{M}$ in scattering processes is not known and can only be expressed as a perturbation series in the coupling parameter of the corresponding quantum field theory, without really worrying about the details of regularization and renormalization techniques. In this chapter, we shall scrutinise these techniques in more detail within the wider context of quantum field theory in order to examine the implications of these strategies to the relationship between the background theory and its models.

The main idea of perturbational analysis is to solve intractable problems by treating them as small adjustments to well known tractable problems. This is achieved by splitting the total Hamiltonian H of the system in question into a free and an interacting part as in

$$H = H_0 + gV \tag{3.2}$$

where H_0 is the free Hamiltonian whose eigenstates we can calculate analytically and gV represents the Hamiltonian of the interacting model with an interaction potential V parametrised by a coupling parameter g. Given that the coupling strength of the theory is weak – as for instance in QED and in high-energy quantum chromodynamics (QCD) – this Type 1 assumption is of course mathematically justifiable since we can safely assume that the solutions of the total Hamiltonian are close enough to the well known solutions given by the free Hamiltonian H_0.

One of the most important quantities to be obtained by applying perturbation theory is the S-matrix, the operator that maps the initial state $|\psi_i\rangle$ of a physical system undergoing a scattering process to the final state $|\psi_f\rangle$, defined by

$$\hat{S}\,|\psi_i\rangle = |\psi_f\rangle \tag{3.3}$$

The importance of the S-matrix to experiments of particle phenomenology boils down to the close relationship of its elements – i.e the Feynman amplitudes $\mathcal{M}$ discussed in detail in Chapter 1 – to the (indirectly) measured scattering cross sections of various processes. The S-matrix thus contains some of the theoretical predictions of the empirically relevant model of the theory upon which the empirical adequacy of the model can be tested. In order to set up an approximating perturbation series for each of the S-matrix elements one first needs to obtain the time evolution operator in the interaction picture in terms of which the S-matrix can be expressed.[5] The transformations between the time evolution operators and states in the Heisenberg picture and the Schrödinger picture allow for the perturbative (iterative) expansion of the time evolution operator $U_I(t, t_0)$ in the interaction picture, defined by $|\psi_I(t)\rangle = U_I(t, t_0)\, |\psi_I(t_0)\rangle$, into a series of the form:

$$U_I(t, t_0) = \sum_n^{\infty} \frac{(-ig)^n}{n!} \int_{t_o}^{t} dt_1 ... \int_{t_0}^{t} dt_n T[V_i(t_1)...V_I(t_n)] \tag{3.4}$$

where the time order product T rearranges the operators in descending order with respect to their time arguments. By plugging the perturbative expansion of the time operator into the definition of the S-matrix with respect to the time evolution operator of the interaction picture

$$S_{if} = \lim_{t_i \to -\infty} \lim_{t_f \to \infty} \langle\psi_i|\, U_I(t_f, t_i)\, |\psi_f\rangle \tag{3.5}$$

we obtain the Dyson expansion of the S-matrix

$$S_{if} = \sum_n^{\infty} \frac{(-ig)^n}{n!} \int_{-\infty}^{\infty} dt_1 ... \int_{-\infty}^{\infty} dt_n T[V_i(t_1)...V_I(t_n)] \tag{3.6}$$

Up until this point the process of constructing an empirically relevant model from the background theory's theoretical models only involves various mathematical approximation techniques that are mathematically justified, such as the reiterative approximation method by which one obtains the Dyson expansion, and the assumption that the solutions of the total Hamiltonian of the system are close to the solutions of the free Hamiltonian. However, when this perturbative method is applied to realistic quantum field theories like QED, the final result is an expression of the Feynman amplitudes $\mathcal{M}$ as a power series of the coupling parameter g which has the following generic form

$$\mathcal{M} = \sum_n^{\infty} g^n \int_{-\infty}^{\infty} dk A_n \tag{3.7}$$

where A_n stands for a multiple integral over momentum k at each n^{th} order term of the series. The main idea is that for those interactions where the coupling constant is relatively small, the first factor g^n vanishes as the series proceeds to higher orders, and thus the contribution of higher order terms decreases significantly. All we need in order

---

[5] The interaction picture is the intermediate picture between the Schrödinger picture, in which states evolve in time under the total Hamiltonian and operators are stationary, and the Heisenberg picture, in which states are stationary and operators evolve under the total Hamiltonian.

to calculate the coefficients of the series for each order (and consequently the S-matrix elements) is therefore to evaluate a set of multiple integrals over momentum space.[6]

This is where three significant problems arise however. The first problem concerns the fact that each individual term of the series often appears to diverge either at very high energies (as $k \to \infty$) or at very low energies (as $k \to 0$) causing the so-called *ultraviolet* and *infrared divergences* respectively. This is the notorious '*problem of infinities*' in quantum field theory and, as we shall see, the solution to this anomaly comes from various regularization and renormalization techniques. This leads to a further complication however, since the various mathematical methods used in these approaches often violate some of the fundamental principles of the relevant background theory, breaking this way the solid deductive nature of the relationship between the background theories and its models. The third problem comes from the fact that the total sum of the series is sometimes either not known to converge or known not to converge, and thus a legitimate question arises here as to whether the series is indeed a good approximation to the quantity in question. Moreover, as one may notice from the form of the Dyson expansion series in Eq. 3.6, the number and complicatedness of each of these integrals increases rapidly for higher orders and thus, in practice, only the sum of the first few terms is evaluated and compared to experimental results, sometimes with great success.

The crucial question here is whether the calculative methods used at this stage can be mathematically justified making the introduced idealizations controllable, as in the simple case of the ideal pendulum. If so, then one is allowed to maintain that the final products, i.e. the empirically relevant models that are eventually put to test by comparison with experimental results, are still derivable from the background theory via the available theoretical models, saving this way van Fraassen's maxim on the importance of the latter. The next section discusses these complications and shows how the only available justification for the introduced approximations is one of a purely *instrumental nature*. As a result, the strong relationship between the final models and the background theory, required by van Fraassen's favourable semantic view, breaks down.

---

[6] As already discussed in Chapter 1, these integrals are often calculated with the help of Feynman diagrams. Eq. (1.5) is essentially a specific instantiation of Eq.(3.7) in which the first order term of the perturbative series is calculated using the corresponding Feynman diagram.

3.3 Breaking the link with the theory

3.3.1 Cutting off our ignorance

The first problem that needs to be addressed in the construction of empirically relevant models in perturbative QFT is the presence of ultraviolet and infrared divergences in the individual terms of the power series in (3.7) which makes the calculation of the coefficients impossible. The elimination of these infinities is achieved by various methods of *regularization* in which a divergent integral A_n is redefined as a function of a new parameter ξ – the regulator – which must satisfy the two following *pragmatic constraints*: (i) finite values of the regulator must render the integral $A_n(\xi)$ finite and (ii) if the regulator is removed by taking its limit to infinity the retrieved result is the original divergent integral A_n.

A standard regularization method is the Pauli-Villars regularization in which a cut-off limit Λ is introduced for the ultraviolet and infrared domains of high and low energies respectively, beyond which the value of the integral is taken to be zero.[7] In the case of ultraviolet divergences, this is equivalent to the introduction of a Type 1 assumption according to which

$$\sum_n \int_\Lambda^\infty A_n \equiv 0 \tag{3.8}$$

The introduction of the cut-off limit Λ should be understood here as a physical construct introduced *by fiat*, representing *our ignorance* about the validity of our methods beyond this limit. It is not a mathematically or physically justified constraint similar to small angle assumption in Eq.3.1 for instance. Rather, the only available justification for the introduction of the cut-off limit is *instrumental*, in that it aims to make the calculations of the series' coefficients possible. In other words, we simply ignore contributions to the integrands for values of momentum greater than Λ in order to remove the infinities and make the integrals physically relevant. A similar technique is used for the removal of the infrared divergences in the low momentum domain that typically occur in theories with massless particles, such as the photons of QED. In these cases, an infrared cut-off limit is introduced for a small but non-zero value of momentum below which the

---

[7] Another popular method of regularization is dimensional regularization where, roughly speaking, the calculations are initially carried out in a d-dimensional spacetime, and the cut-off dependences appear as we take the well-defined limit $d \to 4$. The Pauli-Villars method discussed here has the advantage of being physically more transparent as opposed to dimensional regularization which is considerably more abstract in nature. Some less popular methods of regularization are the lattice regularization, the zeta function regularization and the causal regularization.

contribution of the integrands to the sum is neglected.[8] These methods eventually remove the undesirable infinities from the first few terms and thus the series provides an estimate of the quantity in question – e.g. the S-Matrix element – up to the order for which the infinities have been removed.[9]

A direct implication of regularization is that the approximate value of the scattering amplitudes $\mathcal{M}$ now depends on the arbitrarily chosen value of Λ. However, given that there are no particular reasons for choosing a specific value for Λ over a different one, say Λ^2, this should be worrying; different values of the arbitrarily chosen regulator will simply give different results for the scattering amplitudes, which is of course unacceptable. The scattering amplitudes provided by these models express the empirical content of our theory and thus they should not be a function of the theorists' arbitrarily chosen value for Λ.[10]

The solution to this problem comes from *renormalization*. In this method the unwanted Λ-dependence of the integrals, and consequently of the scattering amplitudes, is eliminated by extracting the value of the coupling parameter g from scattering experiments at a particular energy scale μ. Once the value of the coupling parameter is obtained, the original 'bare coupling' g is replaced by the new renormalized value $g_R(\mu)$ (sometimes called the 'physical coupling parameter') which is now a function of the energy scale μ at which the experiment was performed.

As a concrete example, consider the perturbation series for the scattering amplitude of a meson-meson scattering process up to the second order correction. Once the ultraviolet cut-off is introduced, the amplitude $\mathcal{M}$ becomes a finite 'cut-off dependent' quantity of the form

---

[8] For a philosophical discussion of the treatment of infrared divergences see Miller (2021a).

[9] For textbook style expositions of regularisation and renormalization methods in quantum field theory see Peskin and Schroeder (1995), Zee (2010) and Duncan (2012). For expositions of these methods aiming particularly at philosophers see Butterfield and Bouatta (2015), Wallace (2018) and Williams (2018).

[10] It should be noted here that the choice of Λ is not always completely arbitrary. For instance, when we want to abstract away from the interactions of a heavier, higher-energy particle, Λ is chosen to be well below the mass of that particle and above the mass of the particles we are interested in studying. However, given that the range between the two masses is significantly large, the choice of the exact value for Λ is still, more or less arbitrary. Butterfield and Bouatta (2015, p.448) also point out that in some cases the background theory *hints* towards a range of values for the cut-off limit that are plausible to take. An example of such a suggestion for the cut-off limit comes from the Lamb shift in QED, which suggests the electron's Compton wavelength as a natural lower limit for distance d. Nonetheless, even in these cases no rigorous mathematical or physical justification can be given and thus the final choice of Λ remains arbitrary.

$$\mathcal{M} = -ig + iCg^2[\log\left(\frac{\Lambda^2}{s}\right) + \log\left(\frac{\Lambda^2}{t}\right) + \log\left(\frac{\Lambda^2}{u}\right)] + \mathcal{O}(g^3) \tag{3.9}$$

where C is a numerical constant, and the kinetic variables s, t and u (also known as the Mandelstam variables) are functions of the square of the energy at which the particles are scattered, and are related to rather mundane quantities such as the centre of mass energy and the scattering angle. Once the value of the coupling parameter is experimentally obtained, the bare coupling parameter g in Eq.3.9 is replaced by the renormalized coupling parameter g_R. By carrying out some simple algebraic calculations one is left with a new 'cut-off independent' expression

$$\mathcal{M} = -ig_R + iCg_R^2[\log\left(\frac{s_0}{s}\right) + \log\left(\frac{t_0}{t}\right) + \log\left(\frac{u_0}{u}\right)] + \mathcal{O}(g^3) \tag{3.10}$$

where the values of s_0, t_0 and u_0 are defined by the particular energy at which the coupling parameter g_R was measured. At this stage, the dependence on the cut-off limit Λ is dropped out, however, as one may notice the scattering amplitude now depends on the ratio between the energy scale μ at which the value of $g_R(\mu)$ is measured (captured by s_0, t_0 and u_0) and the energy scale of the future scattering experiments by which the empirically relevant model will be tested (captured by s, t and u).

What is so special about the chosen energy scale μ however? And what if the renormalized parameter $g_R(\mu)$ was measured at a different scale μ'? One might think here that this is not much of an improvement since the dependence on the arbitrarily chosen value of Λ has simply been shifted to a dependence on the energy scales μ at which we are able to conduct experiments and extract the value of the renormalized coupling parameter. The standard way to deal with this situation is to see quantum field theory as a theoretical framework for providing *effective field theories* (like QED and QCD) that only describe physical phenomena at a particular range of energy scales (up to μ) and not as a fundamental theory of physics. This idea became dominant in the late 1960s when a new approach to renormalization, the renormalization group (RG) flow, was developed by the works of Wilson, Fisher, Kadanoff and others.

The renormalization group method provides the mathematical apparatus for a systematic investigation of the changes in the couplings of a theory with respect to the changes in the energy scales. Leaving the technical details aside, the main idea of this modern approach to renormalization is captured by the renormalization group flow equation

$$\mu\frac{\mathrm{d}}{\mathrm{d}\mu}g_i(\mu) = \beta_i(g_i) \tag{3.11}$$

which determines the so-called *beta function* $\beta_i(g_i)$ as the differential change of the coupling parameters $g_i(\mu)$ of a quantum field theory with respect to a small change in energy scale. If the theory happens to have several coupling constants $g_i, i = 1, ..., N$, then the beta function is a function of all of the couplings in the theory $\beta_i(g_1, ...g_N)$. One can think of the renormalization group equation as defining an N-dimensional space of theories in which $(g_1, ..., g_N)$ are the coordinates of a particle that *flows* in the space as the energy scale μ increases. In the high-energy regime where we have no experimental access to measure the values for the coupling parameters, there are three possible behaviours of the theory's couplings depending on the beta function:

(I) The renormalization group flow hits a point (also known as an attractor) at which the beta function becomes zero for all couplings. This is the so-called fixed point $g*$ which basically implies that the theory becomes scale-invariant above a certain energy scale since all couplings converge to a fixed constant value. If a theory demonstrates this behaviour, it is said to be *asymptotically safe.*

(II) The second possibility is a special case of asymptotic safety in which the beta functions are negative for all the couplings in the theory and thus the values of the couplings decrease as the energies become higher and higher until they eventually hit a fixed point which is equal to zero ($g* = 0$). Accordingly, for lower energies (and large distances) the coupling strengths increase rapidly until they become divergent at some low, but finite, scale. This behaviour is thought to be responsible for the phenomenon of quark confinement in QCD, and theories of this kind are said to be *asymptotically free.*

(III) The third possibility is the case in which the beta function is positive for at least one of the couplings in the theory, implying that the coupling increases indefinitely until it blows up to infinity at a very large but finite energy level (μ_{Landau}) at which the theory is said to encounter a *Landau pole.* Such theories are ill-defined and are considered as trivial.[11]

---

[11] It is worth noting that in addition to the first two standard possibilities where the RG flow approaches attractors that are plain isolated fixed points, there is another – less discussed – possibility where the flow exhibits a cyclic or even chaotic behaviour (Mozorov & Niemi, 2003; Curtright, Jin & Zachos, 2012). This possibility was also noted by Wilson himself (1971) and although the existence of such cyclic or chaotic RG flows has not been yet confirmed, several models have been developed as candidates for limit cycle behaviour (Leclair, Román & Sierra, 2003).[I am grateful to my supervisor Karim Thébault for bringing this additional possibility to my attention.]

Unfortunately, this rather attractive theoretical picture is very difficult to implement in practice since, for the time being, there is no way of calculating the beta functions in models of realistic QFTs. At best, what we have available are further approximation techniques for approximating the beta functions of some theories, which further *suggest* a possible behaviour of these theories in high energies. Interestingly, what actually happens here is an attempt to justify the assumptions introduced by regularization and renormalization, by introducing further assumptions, creating this way a form of regress in which idealizations are justified by further idealizations. This is not to say that this is a vicious regress however. Whether the justification of an approximation by means of a further approximation is epistemically safe is an interesting open question and deserves to be studied on its own merit in future work.

What is important for the purposes of our discussion, is whether the final products of renormalization, that is, the empirically relevant models that are eventually put to the test by comparison with experimental data, can still be thought of as products of the background theory, based on the initial non-interactive models of QFT. If one insists in preserving a strong deductive relation between the background theory and its models, one needs to show that each of the aforementioned introduced assumptions is somehow a controllable idealization. However, while this seems to be achievable during the first stages of model construction in perturbative QFT, things become significantly obscure once regularization and renormalization come into play. Given that both methods are mainly motivated by pragmatic criteria such as the two initial constraints for the regulator and the fact that the energy scale of renormalization is essentially defined by our available technology to reach high energies in scattering experiments, it is hard to see how one can still maintain that empirically relevant models in perturbative QFT can be considered as theoretical models in van Fraassen's sense.

This point is also reflected in Alex Franklin's (2020) recent work on the effectiveness of Effective Field Theories. Commenting on Butterfield (2014), Franklin notes that the appeal to the fixed point structure of the RG method in order to justify renormalization is ill suited. His claim is that RG is best seen as a mathematical framework that *codifies* rather than explains the renormalizability of theories in QFT, in that it allows us to mathematically establish whether or not a theory is renormalizable, but nonetheless, it does not provide a sufficient explanation of the physical reasons behind renormalization. Franklin then goes on to suggest that the explanation for the renormalizability of theories is to be found instead in the physical fact that, in renormalizable theories, high-energy couplings have negligible magnitude at lower energies and thus they can be removed from the mathematical treatment in low-energy models since they are irrelevant.

The autonomy of scales in which Franklin appeals to explain renormalization is indeed a necessary and widespread assumption in high-energy physics and is seen by scientists as the best possible explanation behind the effectiveness of the various EFTs. This fact however does not undermine the claims I wish to make here. What is at stake for the purposes of our discussion is whether the empirically relevant models of perturbative QFT are the products of controllable idealizations for which either a mathematical or theoretical justification can be given. The RG method aims in giving a mathematical justification but, as we have seen, in lack of a mathematically rigorous method for calculating the beta functions in realistic models, it currently fails to do so. The autonomy of scales on the other hand, could be seen as providing a possible theoretical justification, however, while this is true, the point is that, in practice, the current formulations of QFT are unable to provide a precise and fundamental theoretical principle based on which a rigorous theoretical justification can be given in order to make the introduced idealizations controllable. This is precisely what causes the breakdown of the relationship between theory and models in perturbative QFT, undermining this way the importance of theoretical models in the context of constructive empiricism. The next subsection discusses two further complications reinforcing this claim.

3.3.2 Haag's theorem and the overall divergence of the series

A further complication in the relationship between the resulting models in perturbative QFT and the background theory comes from Haag's theorem which, roughly speaking, shows that perturbative QFT models rest on an inconsistent set of assumptions. In particular, the theorem shows that given the basic assumptions of the axiomatic formulation of QFT the interaction picture which is necessary for the construction of empirically relevant models in perturbative QFT does not exist. This means that if we wish our theory and models to be mathematically consistent, then free and interacting fields cannot be represented by the same Hilbert space, as the interaction picture requires. The theorem is thus often taken to undermine the validity of empirical claims based on perturbative models (Earman & Fraser, 2006).

Miller (2018) argues extensively that Haag's theorem should not be seen as posing any serious problems to claims about the empirical adequacy of particular quantum field theories. The main reason is that the techniques used in regularisation and renormalization actually render some of the fundamental assumptions required to prove the theorem false. For instance, the Pauli-Villars method violates the unitarity of the time evolution operator and breaks the gauge invariance symmetries of the theory, while the lattice method preserves gauge invariance but violates Poincare invariance which is essential for the proof of the theorem. Hence, the very act of regularizing the theory in order to obtain models from which numerical predictions can be derived and then compared to outcomes of measurements simply renders the assumptions of the original unregularized theory inapplicable to the models. In simple words, just like many phenomenological models in physics, the models of perturbative QFT violate some of the fundamental theoretical principles of their respective QFTs. In light of these issues, it is hard to see how one can maintain the desirable link between the background theory and the resulting models that is necessary for the claim that empirically relevant models are deductively connected to their theoretical counterparts and the background theory. In essence, Haag's theorem implies that either one will obtain theoretical models in perturbative QFT that are non-realistic and thus empirically inadequate, or one will have realistic and empirically adequate models which nonetheless violate some of the fundamental theoretical principles of the background theory.

Finally, one last complication concerns the large order behaviour of the perturbative series as a whole, independently of the offending integrals in the individual terms. Recall that in addition to the fact that the integrals in each term of the series diverge, there is often strong evidence that the sum of the series at all orders diverges as well. Given the infinite number of terms and the increasing intractability of the integrals as one proceeds in higher order terms, what happens in practice is that only the first few terms of the series are usually calculated and compared to experimental results, often by using different techniques for calculating each term of the series. Nevertheless, in several cases the sum of the first few series turns out to be in tremendous agreement with experimental results leaving physicists and philosophers wondering how the empirical success of the truncated series should be explained.[12]

Miller (2021b) notes that a possible explanation of the empirical success of the first few terms of the series can be given in terms of *strong asymptoticity*.[13] Unlike convergent series, when a divergent asymptotic series is used for approximating a function, the sum of the first few terms is considerably close to the exact value of the function. However, as one includes more and more terms to the sum, the value of the series increasingly

---

[12] The theoretical prediction for the anomalous magnetic moment of the electron, discussed in first chapter is a nice example of this practice. The current state of the art only allows the calculation of the first five terms which requires – amongst other things – an evaluation of 12672 Feynman diagrams in the tenth-order perturbation theory (Kinoshita 2014).

[13] As Miller notes, this explanation was also given by Freeman Dyson (1952) himself in his arguments for the divergence of the perturbative series.

diverges until it becomes infinitely different. Hence, if the perturbative series for, say, the matrix elements of a meson-meson scattering process is asymptotic to the unknown exact solution, it should be no surprise that the sum of the first few terms is often found to be in agreement with experimental results.

The problem however, is that this explanation is currently a mere conjecture. The condition of strong asymptoticity showing that a perturbative expansion uniquely satisfies an exact solution is only satisfied if a series can be shown to be Borel summable.[14] However, just as with beta functions, Borel summability has not been rigorously shown to hold in any phenomenologically interesting models of quantum field theory. Duncan's (2012, p.403) comments on the difficulty of proving Borel summability in models of perturbative QFT are rather illuminating: 'the property of Borel summability is an extremely fragile one, and one which we can *hardly ever* expect to be present in interesting relativistic field theories' (emphasis added). Hence, in lack of a rigorous mathematical justification, the possibe explanation of the empirical success of the perturbative series in terms of strong asymptoticity is not sufficient to render the involved idealizations controllable. This fact provides further support to the claim that the relationship between the empirically relevant models in perturbative QFT and the background theory is not deductive, and cannot be described in terms of controllable idealisations.

3.4 Models of perturbative QFT are not theoretical

In light of the above observations, it should be clear by now that the empirically relevant models of perturbative quantum field theory are not theoretical in van Fraassen's sense. The mathematical idealizations involved in these models via regularization and renormalization are far from being controllable and in most cases the only available justification for the introduction of Type 1 assumptions in these models is merely instrumental. Hence, one might say at this point that the proponents of the semantic view are confronted with a two-horned dilemma regarding the exact meaning of theoretical models and their place in modern physics. One option is to relax the criteria for what counts as a theoretical model and say that even in cases where the empirically relevant models are constructed by introducing non-controllable idealizations, they are

---

[14] In short, Borel summation is a mathematical method that is particularly useful for summing divergent asymptotic series. Suppose a formal power series $\sum_{k=0}^{\infty} \alpha_k \lambda^k$ of a function $f(\lambda)$ and define the *Borel transform* B to be its equivalent exponential series $B(\lambda t) \equiv \sum_{k=0}^{\infty} \frac{dk}{k!} t^k$. If the Borel transform can be used to produce a unique reconstruction of $f(\lambda)$ then the series is Borel summable. For a more rigorous discussion of Borel summation in the context of perturbative QFT see Duncan (2012, pp.400-403). See also Miller (2021) for a philosophical discussion.

still models of the theory from which they are inspired. Indeed, this is the language that physicists seem to use when referring to theoretical models in a much more loose sense than what the semantic view implies. This, however, becomes a trivial claim about physics which merely says that all empirically adequate models are loosely related to a background theory, losing this way the strong deductive relationship required by van Fraassen's favourable semantic view of scientific theories. The other option, is to abandon the overall importance of theoretical models and concede that many empirically successful models in physics are indeed not theoretical. However, given van Fraassen's strong views on the importance of theoretical models within the context of the semantic view, it is unlikely that he would go for the latter option.

How can one get out of this situation then? A standard way of dealing with non-theoretical models within the semantic view of scientific theories is to dismiss their importance and see them as temporary constructs of preliminary physics (da Costa & French 2003, pp.55-56). While this might be a good way of understanding the nature of Redhead's floating models, it is not an available option in the case of perturbative QFT. The perturbative approach to QFT has been pivotal in the 'triumph' of QED which begun in the 1930s and culminated in 1947 with the measurements of the Lamb shift (Lamb & Rutherford, 1947) and the anomalous magnetic moment of the electron (Kusch & Foley, 1948) based on calculations in lowest-order perturbation theory and renormalization.[15] Following the huge success of QED, perturbative QFT still remains a highly successful approximation method for constructing models in scattering experiments and to say that such models are temporary or part of some sort of preliminary physics is to dismiss a large part of contemporary scientific practice.

Van Fraassen's approach stands on a better ground. Referring to phenomenological models, his verdict is that the appearance of highly successful non-theoretical models often initiates 'a revolutionary change' of theories under the pressure of 'new phenomena', where new phenomena can be understood as appearing through experimental data that cannot be explained or fit into the existing theoretical models of the background theory (2008, p.310-11). In this context, one might think of the models in perturbative QFT as having a similar role to successful phenomenological models as vital parts of a transitional stage in physics calling for further work which will eventually lead to theoretical interacting models of scattering processes. This is indeed reflected by the fact that QED and QCD are currently seen as effective field theories that are basically good approximations for the low energy regime of a final 'theory of everything' that applies to

---

[15] See Huang (2013) for a critical historical approach on the role of renormalization in QED.

all energy scales, such as string theory. There is no doubt that both these views represent a significant part of the scientific practice, however, the claim I wish to make here is that these are not the only possible ways to understand the existence and importance of non-theoretical models in physics.

Whether van Fraassen is right or not depends on how one understands his claim. If it is seen as a descriptive claim about how physics *actually* makes progress, then, given that for the last fifty years or so there has not been yet a successful axiomatic reformulation of the theoretical framework of QFT in order to produce interacting models, the description is inaccurate. If it is seen as a normative claim about how physics *should* make progress, then, given the fact that perturbative QFT models are seen as parts of an effective field theory, the description is correct. Insofar as empirically successful models are not derivable from a well defined theory, there will always be a constant search for a more fundamental framework in which these models can be either embedded or explained as limiting approximations. This is in accordance with the view, shared by many philosophers, that the existence of four-dimensional interacting models satisfying the fundamental principles of QFT is a necessary condition for considering a particular QFT as a respectable foundational theory.

An alternative line of thought that is worth considering here, comes from James Fraser's recent work (2020) on perturbation theory. Drawing on Norton's (2012) distinction between approximations and models, Fraser proposes a view that eliminates models entirely from the overall picture of perturbative QFT. In this context, approximations are understood as methods for producing functions that inaccurately describe a target system's properties, without necessarily resembling that system. Models, on the other hand, are understood as *structural characterisations* of a physical system that carry some sort of representational capacity and whose dynamic behaviour is defined by their state space. Fraser's ultimate claim is that it is difficult to understand the output of the standard perturbative formalism as a model of a physical system: it does not specify the space of states, nor the dynamics for interacting systems such as the scattering of two mesons for instance. It merely gives an approximation of a physical quantity such as the cross-section of a scattering process. Based on these observations, he then makes the further claim that this is a problem for the constructive empiricist:

> The constructive empiricist states their epistemic commitments with respect to models, taking them to be empirically adequate rather than representationally faithful. If I am right then perturbative QFT cannot be read in these

> terms [...]: it does not provide us with physical models at all, empirically adequate or not (*ibid.*, p.19-20).

Given that one endorses Fraser's terminology of what counts as a model of a physical system, his claim that perturbative QFT does not produce models is legitimate. As opposed to other more traditional methods of modelling whereby a full description of the physical system is given (e.g. the ideal pendulum), the formalism of perturbative QFT is indeed considerably more abstract in nature. It is therefore hard to see how these methods can be seen as producing any sort of structural characterisations of a physical system with a certain dynamic behaviour determined by the state space of the model. Whether this poses a serious problem for the constructive empiricist however, is a different matter.

Although not made explicit in the text, what Fraser seems to have in mind when he makes this claim, is van Fraassen's 'state-space approach' (1980, 1989), where models are understood as picking out a physical system's evolution in time by forming trajectories in state spaces. However, if one adopts a broader understanding of models, such as the one presented in Chapter 1, then clearly, the perturbative approach to QFT produces models whose mathematical framework contains the approximating function which with an appropriate interpretation is linked to a measurable quantity such as a cross-section. What is more, given van Fraassen's definition of a theoretical model in Section 3.1, a theoretical model need not be a structural characterization of a physical system whose dynamic behaviour is represented by a temporal trajectory in the space of states. Rather, what defines a theoretical model is the fact that it specifies a family of observables for a physical system related by a set of equations and, most importantly, that it provides the stochastic response function for the observables in question which eventually encapsulates the model's prediction to be tested by experiments. This stochastic response function is part of the mathematical framework of a model, ans is precisely what Fraser describes as an approximation function for the S-matrix element of a scattering process. Moreover, the state space of the model does not necessarily resemble the dynamic behaviour of the physical system – like the phase space in classical mechanics for example – rather, it should be understood as an abstract mathematical artefact capturing the mathematical interdependence of the various variables of a model.

Fraser is therefore right when he says that the formalism of perturbative QFT does not produce structural characterisations of physical systems with a specified dynamic behaviour resembling the actual behaviour of their targets. However, perturbative QFT does produce models, insofar as modelling is understood in a broader sense as the general practice of constructing mathematically tractable structures for the prediction and calculation of the physical quantities of a given system based on a physical theory. This view is undoubtedly closer to van Fraasen's description of theoretical models and – arguably – the scientists' overall attitude towards the concept of the model of a physical system or process. Hence, the fact that perturbative QFT does not produce structural characterisations of physical systems does not really pose any serious problems for the epistemic commitments of the constructive empiricist with respect to the empirical adequacy of the models. The real problem for the constructive empiricist – if it is a problem after all – is that these models are not theoretical.

As already noted, van Fraasen's emphasis on the importance of theoretical models in his constructive empiricist programme, and in science in general, is a result of his strong inclination to the semantic view of scientific theories. One may still advocate an empiricist approach to science however, without necessarily endorsing the semantic view and its implications on the importance of theoretical models. The epistemic commitments of an empiricist can be stated both in terms of the theoretical models of a physical theory *and* the various non-theoretical models that are occasionally developed by physicists. The only difference is that in the first case, the empirical adequacy of the models can be attributed to the background theory as well, whereas in the latter case the empirical adequacy concerns primarily the empirically relevant (non-theoretical) models, leaving the status of the empirical adequacy of the supporting theory unsettled.

3.5 Conclusions

The fact that perturbative QFT produces non-theoretical models as a result of non-controllable idealizations raises a very interesting question regarding the empirical adequacy of perturbative QFTs. Does the empirical adequacy of the non-theoretical models of perturbative methods guarantee the empirical adequacy of the background theory as well? The suggestion here is to see the empirical adequacy of the non-theoretical models as providing supportive – but not conclusive – evidence for the empirical adequacy of the background theory (at least within the energy scales in which these models provide accurate predictions). However, what is important is not whether the answer to this question is positive or negative; after all this might turn out to be a matter of the definition of the empirical adequacy of a theory. What really matters, are the reasons why we should or should not attribute the empirical adequacy of non-theoretical models to their supporting theories. By highlighting the fact that theory and models are not clearly connected in perturbative QFT the aim here is to pave the way for further work on these issues.

Moreover, with respect to our discussion in Chapter 1 about the role of models in the wider aims of physics and science in general, we have already highlighted the contribution of Feynman diagrams in QFT towards the economy of thought. In addition, the present chapter nicely illustrates the important contribution of models for extracting predictions from a background theory. The masterful construction of models in perturbative QFT via the techniques of regularisation and renormalisation is a prime example of how the art of modelling in physics facilitates the derivation of empirical predictions, especially in the particularly interesting cases whereby a clear axiomatic formulation of the theory (as in axiomatic QFT) fails to produce any physically relevant models. Finally, with respect to the aim of explanation, the lack of a rigorous mathematical justification for the introduced idealisations in the models of perturbative QFT, combined with Fraser's (2020) observations about the lack of a structural characterisation of a physical system in these models, perhaps implies that the explanatory status of these models is – at least – unclear. The non-theoretical models of perturbative QFT could thus be understood as an example of an important type of models in physics which fail to provide a satisfactory explanation of the underlying physical phenomena, but surely, this is another very interesting topic that deserves to be explored further on its own merits.

In all, the study of the modelling techniques of perturbative quantum field theory, has shown that modelling via regularization and renormalization produces non-theoretical models that are significantly different than phenomenological models, indicating the presence of a third – understudied – type of models in physics. Given the importance of renormalization techniques in the methodology of modern physics, the further claim here is that the presence of these models significantly undermines both van Fraassen's maxim on the importance of theoretical models in physics and the semantic view of scientific theories in general. This conclusion forces us to reconsider the relationship between theories and models as well as the exact role of theoretical and non-theoretical models in scientific practice.

CHAPTER 4

WHAT IS A DATA MODEL?

AN ANATOMY OF DATA ANALYSIS IN HIGH-ENERGY PHYSICS

So far, we have been mainly concerned with the relationship between models and their background theories, and the ways in which the former can be deduced from the latter. In this chapter, we turn our attention to the relationship of models with the experimental data that typically appear in their refined form as *data models*. The constantly growing integration of science and technology during the last decades has brought science in the new 'era of big data'. Modern experimental setups and other advanced methods of data collection often result in enormous datasets calling for more and more sophisticated methods of data analysis in order to enable the comparison of the experimental results with theoretical hypotheses. The well known Hypothetico-Deductive method whereby theoretical hypotheses are reinforced – or, in Popperian terms, corroborated – in the light of new data nicely captures a large part of the scientific practice, however, at the same time it provides an oversimplified and unrealistic picture in which important details are left aside. How exactly are theoretical hypotheses eventually confronted and tested by experimental results given that the latter are often produced in the form of large datasets and in a language that is not accessible to the theory?

Patrick Suppes (1962) answered this question by pointing out that theoretical hypotheses are not directly confronted with the raw unprocessed data from experiments, rather, they are only confronted with *models of data*. What exactly is a model of data however? Suppes explains in an earlier work: 'The maddeningly diverse and complex experience which constitutes an experiment is not the entity which is directly compared with a model of a theory. Drastic assumptions of all sorts are made in reducing the experimental experience [...] to *a simple entity* ready for comparison with a model of the theory' (1960, p.297, emphasis added). In Suppes' mind, a data model is a simple entity that incorporates what is often a very complicated and sophisticated experimental process, into a simple result which is eventually compared with the theoretical predictions of a theory or a model.

While Suppes' remarks on data models sowed the seeds for further significant work on the philosophy of data, the interest of philosophers of science in data was mainly revived by the seminal works of Bogen and Woodward (1988), Woodward (1989) and van Fraassen (1980, 1989). In response to van Fraassen's well-known view that the empirical adequacy of a theory is measured by its ability to save the observable phenomena, Bogen and Woodward emphasized the distinction between data and phenomena and argued that theory often saves the non-observable phenomena, rather than the observed data. Within this context, a large part of the discussion that followed on the philosophy of data (e.g. McAllister 1997; Glymour 2000; Harris 2003; and Massimi 2007) has mainly focused on the relationship of data with the physical phenomena that they are often taken to represent.

In the more recent literature on data, this tendency of philosophers to examine the nature of data with respect to the underlying phenomena they represent, has been replaced by a new tendency to closely examine examples of actual scientific practice, in order to explore the methodology of data processing in various scientific fields and the role of data models within them. This approach is most evident in the works of Sabina Leonelli (2015, 2016, 2019) in biology and Alisa Bokulich (2018, 2020) in palaeontology which, as one might expect, are highly influenced by works of non-philosophers on data processing, such as Edwards' (2010) book on climate science.

This chapter follows this recent tendency and aims in expanding the existing literature on the methodology of data analysis into the field of modern High-Energy Physics (HEP). Given that modern large-scale HEP experiments rely on the production of large volumes of data more than any other scientific field, it is surprising that not much has been said about the methodology of data analysis in this field. As Bokulich notes, although many philosophers have followed Suppes in highlighting the importance of data models in science, 'most [of them] have largely black-boxed how data models are produced' (2020, p.794), and this includes the discussions on the philosophy of HEP. The primary aim of this chapter is to fill this gap by closely examining the methodology behind the production of data models in HEP in order to facilitate our understanding of the nature and the role of data models in this field.

Massimi's (2007) work is one of the few exceptions in which a detailed attention to the analysis of data in HEP is given.[1] Defending the thesis that data provide evidence for unobservable phenomena in HEP, Massimi follows van Fraassen – although she argues against him – in providing a logical reconstruction of data models as partially ordered sets. Using the example of the discovery of the J/ψ particle in the 1970's, her analysis focuses on the presence of unobservable phenomena as outputs of data emerging from different contexts, and the ways in which these phenomena constrain the expansion of the theoretical framework to include the possibility of new kinds of entities and new theoretical quantities at a fundamental level. One of Massimi's central claims is that 'often in science there is a fairly long chain between data and the final parameter that the data model is meant to measure' (*ibid.*, p.246). And even though the raw data of the experiment that are essentially the inputs of the data models are observable phenomena – e.g. sparks in chambers and Cherenkov ring images – the output of these models can be interpreted as a manifestation of a new unobservable phenomenon, such as the presence of a new particle. Our goal in this chapter, is to extend this work by further analysing the implications of Massimi's important observation regarding the 'long chain' between the production of data at the early stages of an experiment and the final data model to be compared with the theoretical hypothesis under investigation.

In particular, this chapter explores the nature of data models and their place in HEP by providing a detailed case study of experimental tests of Lepton Flavour Universality (LFU) at the LHCb experiment at CERN. The adopted methodology is characterised by a systematic study of real scientific practice, and falls within the emerging framework of Philosophy of Science in Practice which aims in producing 'productive interactions between philosophical analyses and the study of actual scientific practices' (Ankeny et al., 2011, p.305). By taking a close look at the work of the experimental physicists of the LHCb collaboration, the idea is to depart from the usual theoretical approach of philosophy of science and re-examine the concept of data models, as well as other related questions, strictly in terms of scientific practice. The ultimate goal is to gain important insights to the question of what a data model is by examining the process by which a data model is constructed in HEP experiments. This thorough examination will nicely demonstrate the way in which theoretical hypotheses are eventually connected with experimental results via data models, and will highlight the importance of considerations

---

[1] Some further work on the methodology of data processing in HEP comes from Karaca (2018) and will be discussed in the next section.

regarding the selection criteria, efficiency calculations, data fitting, and uncertainties in the process of constructing a data model that can be compared to a theoretical prediction in HEP. Contrary to the traditional understanding of data models as idealised versions of the raw data perceived by our immediate observations, the proposed understanding of data models does not rely on the problematic distinction between raw and processed data, nor does it involve the process of immediate observation.

Our analysis in the present chapter will also illustrate some of the central ideas discussed at the beginning of the thesis in Chapter 1. As we shall see, the employment of various experimental models in the design and data acquisition stages of the experiment depends on the interpretation of the mathematical framework of the models, since the way in which the parameters of a model and their mathematical relationships are eventually linked with physical quantities of the material of the detectors essentially determines the collection of data for the purposes of the experiment. Moreover, the whole process of deriving a theoretical hypothesis from the standard model and the possible agreement or disagreement with the experimental results, is a clear example of how the use of models facilitates the predictions of a theoretical framework and leads to the construction of new phenomenological models which in turn lead to further refinements or extensions of the theory.

The philosophical lessons we can take from the exposition of the particular case study at the LHCb are abundant and are outlined throughout this chapter. There are however, four main lessons that can be taken to apply in any large scale HEP experiment. The first lesson concerns what Bokulich (*ibid.*) calls *the folk view of data* which, amongst other things, claims that the tampering of data results in their corruption and the decrease of their epistemic reliability. Contra to this seemingly popular conception, it will be shown that not only the epistemic reliability of data often increases via their processing, but also that raw data – i.e. data that comes out of the detector at the early stages of the experiment – are actually useless as they are for the comparison of the experimental results with the theory. One of Bokulich's central aims is 'to make plausible the prima facie *counterintuitive claim* that model-filtered data can – in some instances – be more accurate and reliable than so called raw data, and hence beneficially serve the epistemic aims of science' (*ibid.* p.10, emphasis added). The close examination of our case study in HEP shows that this view – i.e. that model-filtered data are epistemically superior from the raw data obtained from the experiments – not only is not counterintuitive when it comes to the experimental practice of HEP, but rather, it is the *norm* for conducting successful experiments. As will become evident in the following sections, the very nature of experimental HEP makes the interpretation of raw, unprocessed data impossible, and hence, the only way to achieve progress in the field is by collecting and analysing large volumes of processed data.

This also suggests that a clear distinction between raw data and processed data cannot be applied in the context of large-scale HEP experiments. The close examination of the case study on LFU tests illustrates that an understanding of the concept of 'raw data' as data perceived directly from our experience is largely irrelevant to the scientific practice in HEP and what is labelled as raw data and processed data is often merely a matter of convention. Instead of placing data into two distinct categories as raw or processed, what best describes the current scientific practice in HEP is the placement of data in a continuous spectrum in which some datasets are more processed than others, without really worrying where to draw the line between raw and processed data.

The second lesson concerns the further distinction between real and simulated data. The careful scrutiny of the experimental practice at the LHCb illustrates that the boundaries between these two types of data are not as sharp as it is often implied in the literature and not particularly important for the completion of successful experiments in HEP. As we shall see, the final datasets that reach the hands of theoretical and experimental physicists for interpretation are essentially consisted of a mixture of real and simulated data that cannot be distinguished, due to the fact that simulated data are often embedded in real measurement outcomes during the various stages of the experiment. The final data model that is eventually compared to the theoretical hypothesis and provides the 'window' through which theory makes contact with the real world, is essentially a *co-production* of real and simulated data.

The third lesson concerns the various levels in which theory guides the overall construction of the experiment as well as the extraction, interpretation and the further analysis of the acquired data. In particular, the example of the experimental tests of LFU in rare B-decays nicely illustrates that theory-ladenness emerges at three different levels throughout the experiment depending on the scope and the purpose for which background theoretical assumptions guide the experimental process. The possibility of a vicious circularity due to the theory-ladenness of observation will not be discussed in detail here since this is a well-known problem which has already been thoroughly examined by many authors (e.g. Franklin et al. 1989; Brown 1993; Brewer & Lambert 2001; Schindler 2013; Franklin 2015; Beauchemin 2017; Ritson & Staley 2020; Elder 2022). The consensus in these discussions is that theory-ladenness is not necessarily vicious and does not lead to a relativist account of contemporary science. In accordance with this view, our case study shows that the various potential threats of circularity are indeed mitigated by the practice of 'blind analysis' and the implementation of uncertainties in the final result. The focus here will therefore remain on the different levels and the extend to which various theoretical assumptions affect the physicists' decisions in triggering data and their overall understanding of the events at the LHC.

Finally, the fourth and most general lesson concerns the overall process in HEP experiments for the construction of a data model to be compared with the theory. In his seminal paper on data models, Suppes defines models of data 'in terms of possible realizations of the data' (1962, p.253) in the same way that the models of the theory are possible realizations of the theory in the logician's sense. As we have already discussed in Chapter 1, this formal characterization of data models by Suppes closely follows his favourite semantic view of theories which sees theoretical models as set-theoretical structures that are deductively derivable from theoretical sentences. However, as we shall see, the process of building a data model in HEP via the four stages of selection criteria, efficiency calculations, data fitting and uncertainties is way more complicated and less easily formalised than Suppes' discussion would lead one to believe.

Although the final data model is indeed 'a simple entity' as Suppes pointed out, the process of converting the initial data from the detectors into a concise and polished final result in the form of a statistical hypothesis based on the available data is, as we shall see, anything but simple. The complexity of this process mainly stems from the fact that the aforementioned stages do not follow a clear chronological order and cannot always be easily distinguished. Rather, they describe the essential procedures of a long and reiterative process during which data from the experiment are processed and analysed in a number of various ways, including their fusion with data from simulations and the use of highly sophisticated techniques of statistical analysis. During this long process, theory infiltrates the analysis of data at various levels, having clear effects both on the nature of the collected data and their final interpretation. The fact that different theoretical considerations and different techniques of statistical analysis can, in principle, provide slightly different results makes the description of data models as possible realizations of data in a logical sense seem somewhat unsuitable in the context of HEP. Nevertheless, the three types of models in terms of which Suppes described the connection of theory and experiments in his hierarchy of models account – i.e. models of theory, models of experiment and models of data – are useful concepts, and will be used in what follows for sculpting the overall framework of experimental practice in HEP. At the same time, it will be shown that a less stringent version of Suppes' hierarchy of models account is indeed reflected at the practical level in HEP experiments, despite the criticism that has occasionally received.

The structure of this chapter is as follows. Section 4.1 opens the discussion with a defence of Suppes' hierarchy of models account, which will serve as the basis for the present account on the nature of data models and their relationship with theory. In Section 4.2 the focus will be shifted to the necessary theoretical framework for understanding the B-anomalies in particle physics and their usage in tests of the theoretical principle of Lepton Flavour Universality. Section 4.3 will follow with a presentation of the data processing system at the LHCb experiment at CERN, illustrating how theory enters the collection and analysis of data in three different levels. In Section 4.4 the process of constructing the data model representing the final experimental result will be described in four stages. Section 4.5 will follow with a discussion of the distinctions between raw/processed data and real/simulated data. Finally, Section 4.6 provides a brief evaluation of the final data model and its comparison to the corresponding theoretical prediction from the standard model. Section 4.7 concludes the discussion by drawing together the philosophical insights from the examination of the case study at the LHCb.

4.1 Stretching the Hierarchy of Models account

Suppes (1962) begins his analysis of the relationship between theory and experiment by noting that the theoretical principles to be tested do not usually have a direct observable counterpart in the experimental data. Instead, this gap between theoretical predictions and experimental results is filled by a number of different types of models and theoretical principles which Suppes classifies in five levels. At the top level are the models of the theory relevant to the experiment. The main function of these models is to narrow down the typically broad scope of the theory in question into a simple hypothesis H_0 to be tested by the experiment. At the next level one finds the models of experiment. These are models that are 'closer to the actual situation' and whose aim is to adjust the theoretical model to the specific features of the particular experimental setup by providing all the necessary details of how the experiment must be designed and how the data can be linked to the hypothesis in question. At the third level, models of data enter. Suppes describes these as the possible realizations of the data that are 'designed to incorporate all the information about the experiment which can be used in statistical tests of the adequacy of the theory' (*ibid.*, p.258). Finally, in the lowest two levels, are the theory of experimental design which deals with various problems of the experiment that are beyond the particular theory being tested, and what Suppes calls the 'ceteris paribus conditions' which concern every other 'intuitive consideration' of the experimental setup that does not involve formal applications of the theory (e.g. safety rules, control of external disturbances etc.).

Suppes' account has been further elaborated by Deborah Mayo (1996, Ch.5). Although significantly richer in details, Mayo's account maintains Suppes' main idea: theory becomes testable through the models of the theory which provide a distinct primary question or hypothesis to be tested, and experimental results are linked to this hypothesis as models of data. The connection between these two types of models is mediated by the experimental model: 'If the primary question is to test some hypothesis H, the job of the experimental model is to say, possibly with the aid of other auxiliary hypotheses, what is expected or entailed by the truth of H with respect to the kind of experiment of interest' (*ibid.*, p.133). For Mayo, the two key functions of the experimental model are (i) to provide an experimental analogue of the primary theoretical model and (ii) to specify the necessary techniques for linking the experimental data to the questions of the experimental model.

What is also common in Suppes' and Mayo's approach is their emphasis on the importance of statistical and other formal methods of analysis in the construction of data models, as a necessary tool for the successful transition from the level of the theory to the level of the experiment. Suppes conclusion is that once the experimental results are condensed into a simple data model, 'every question of systematic evaluation that arises is a formal one' (Suppes 1962, p.260-1), implying this way that data models are necessarily statistical models, or at least, subject to statistical and mathematical analysis. Drawing on Suppes' empahsis on statistical methods, Leonelli (2019, p.22) has recently criticised Suppes' account by identifying three problems. First, Leonelli notes that Suppes' analysis only deals with numerical data, neglecting the fact that there are also cases where data are not quantitative objects and thus are not amenable to statistical analysis. Second, it is hard to see how Suppes' analysis can be applied in situations of exploratory experiments where the research question under investigation is not clearly stated and thus, it cannot be easily compared with the data model. Finally, Suppes' approach presupposes, according to Leonelli, the ability of researchers to identify what constitutes 'raw data' in the experiment, and overlooks the close connection between the activities of data acquisition and data manipulation.

Leonelli's first observation is correct and lends further support to the claim made earlier that the diversity and complexity of data analysis in various scientific fields makes it impossible to come up with a universal philosophical description for the relationship of theories and data in science. The remaining two observations however, are subject to further analysis. Leonelli's point with respect to the application of Suppes' framework on exploratory experiments stems from Suppes' dictum that the theoretical predictions of a theory are typically expressed in the form of an initial hypothesis and are eventually compared with data models. If there is no theoretical hypothesis to be validated via its comparison with a data model, then Suppes' description is inadequate.

This would be true however, only in the unrealistic cases where scientists are blindly looking for new physics in collider experiments from an Archimedean point of view, independently of any sort of background theory. This is hardly the case in large-scale HEP experiments. The description of the methodology of data acquisition and data processing in Sections 4.3 and 4.4, clearly shows how the very act of collecting and analysing data in HEP experiments is simply impossible without the presupposition of a clear theoretical hypothesis with respect to which the data models are built. What constitutes an exploratory experiment in HEP is not the fact that the research question is not clearly stated, rather, it is the fact that the question is not part of an already well-formed and established theory to be tested.[2] Karaca (2017) also notes that the exploratory nature of HEP experiments concerns the ability of an experiment to achieve a variety of possible outcomes, which as we shall see, can be made possible by the systematic variation of the various experimental parameters.

The example of LFU tests to be used as a case study here, is a clear example of a non-exploratory experiment in which the theoretical prediction of the Standard Model for the R_K ratio is put to the test by building a corresponding data model. However, one might think of a hypothetical situation where either (i) the existing theoretical framework does not provide a precise numerical value of the ratio, or (ii) several competing and not-well established models offer different values of the ratio. In this case, the research hypothesis shifts from 'Is the experimental value of the R_K close enough to the theoretical prediction of the Standard Model?' to the more exploratory question 'What is the value of the R_K ratio?'. Nevertheless, in both cases, the final data model is built with respect to a corresponding theoretical question since the ultimate aim is to fit the data into an already existing or a future theoretical framework. If the data is not in a comparable

---

[2] This approach is also compatible with Steinle's account in which exploratory experimentation 'is driven by the elementary desire to obtain empirical regularities' and 'despite its independence from specific theories, the experimental activity may well be highly systematic and driven by typical guidelines' (1997, p.70).

form with a theory, then this task cannot be accomplished.[3]

Leonelli's third objection is very similar to an objection raised by Karaca (2018), regarding the lack of a modelling concept for the data acquisition process in Suppes' account. Using the example of the ATLAS experiment at CERN, Karaca notes that both Suppes' and Mayo's descriptions leave out a significant aspect of the overall process of bringing together theory and experiment in HEP, which is the specification and organization of the necessary experimental procedures in order to select the required data. This is achieved, according to Karaca, via a model of data acquisition whose key function is to specify the operational and technical details during the data acquisition process and to determine the necessary selection criteria for the rejection of non-interesting events in the LHC collision experiments. While Karaca is right to point out that Suppes' description does not explicitly address the process of data acquisition in HEP experiments, the modified version of the hierarchy of models account that I wish to provide here includes these and other related models within the broad concept of experimental models.

Leonelli's additional point to the discussion is that Suppes presupposes a problematic distinction between the 'raw data' that constitute the 'simple datasets' to be processed and the data models that are eventually compared with the theory. Leonelli draws on Harris' (2003) accurate observation that very often the data that are traditionally referred to as 'raw' are in fact data models, and thus, it is not clear how these models can be compared with theoretical hypotheses. However, this confusion comes from a subtle point regarding Suppes' claims. Suppes definition of data models with respect to their ability to be compared with the theory only applies to the final simple entity which eventually puts the theory to the test. However, Suppes is not saying – or, at least, should not be understood as saying – that *any* data model must necessarily be comparable with a theoretical hypothesis as Leonelli seems to imply. Rather, what Suppes' is saying is that, when it comes to the comparison of theory with experimental results, the entities with which theoretical hypotheses are eventually compared are *necessarily* data models that are subject to statistical analysis. This is a subtle point, but nonetheless it is important for making sense of the fact that very often the various datasets throughout the process of data acquisition and analysis are indeed consisted of – what I shall call – *auxiliary data models*, and whose function is to facilitate the construction of the final data model to be compared with the theory. Moreover, as we shall see in Section 4.5, the

---

[3] This view is also nicely supported by Bokulich and Parker (2021) in a recent paper on what they call the 'pragmatic-representational view of data'. By using an example from climate science, Bokulich and Parker highlight the fact that data and data models are representations that should be evaluated in terms of their adequacy for a particular purpose, in which case is the specific research question.

distinction between raw and processed data is indeed not so clear, as Leonelli and Harris have pointed out, however, it is precisely for this reason that it is also not necessary for describing the scientific practice in HEP.

The deeper lesson to be learned here is that the process of theory testing via experiments in HEP is simply too complicated to be fully captured by a sharp tripartite distinction of three types of models. The modified account I wish to present in the following sections is partially a reconciliation of Suppes' approach with Karaca's remarks, which focuses on three different types of models that constitute a research project. There are however three important caveats to keep in mind. First, the concept of experimental models is significantly extended in order to include *every possible* model related activity which facilitates the connection between the models of theory and the models of data. Second, although it is possible to provide a relatively clear definition (or description) for models of theory and models of data, when it comes to the various types of experimental models the boundaries between them and the two aforementioned types cannot be sharply distinguished and whether one wishes to include a specific modelling activity (such as the models for the specification of selection criteria) in one level or another is up to a certain extend a matter of personal choice. Moreover, for a given collision experiment in HEP, each one of these three types does not consist of a single entity; rather, it should be understood as a cluster of models with similar features serving a common goal.

The third and most important caveat is that the definition for the models of theory in this chapter, is slightly different from the definition of *theoretical models* in Chapters 1 and 3. Recall that in the previous chapters, we followed the terminology of the semantic view and defined theoretical models as mathematical structures that are deductively derivable from a well-established background theory, modulo some controllable idealizations. While Suppes' understanding of the models of theory is undoubtedly closely connected to the semantic view, in this chapter, we shall use the term models of theories in a broad sense to include all types of models we discussed so far, i.e. theoretical models, phenomenological models and the models of perturbative QFT. There are two reasons for this. The first reason is that, on the practical level, the theoretical hypotheses to be tested via the comparison to data models in HEP typically come from all three types of models we have discussed so far. The second reason is that, as we have seen, the intrinsic nature of these models is quite similar, and even though they might be constructed for different purposes, they can all provide theoretical predictions based on the interpretation of their mathematical frameworks. Both phenomenological models and the models of perturbative QFT are also somehow related to background theories, and the only difference is that in the case of a phenomenological model, the background theory might not be a well established theory.[4] Our focus here is therefore, not on the ways in which a model is related to a theory, but rather, on the fact that a model of a theory provides a theoretical hypothesis via its mathematical framework which is tested by a comparison to a data model.

The three different types of models will thus be understood as follows:

- *Models of theory*: A model of a theory is a mathematical tool whose aim is to narrow down the scope of the background theory by providing an experimentally testable hypothesis H_0 (or a number of hypotheses) concerning a specific type of physical processes or phenomena. The background theory providing the hypothesis need not be a well-established and empirically well-confirmed theory. It can also be an isolated and preliminary theoretical framework based on a small class of observations, which would give rise to a phenomenological model. Depending on the nature of the background theory and the model, the hypothesis may concern the exact numerical value of a theoretical parameter (e.g. the fine structure constant at a given energy level, the electron magnetic moment etc.), an estimate of a model parameter in the form of a probability distribution corresponding to a physical quantity (e.g. particle properties such as mass), or a specific relation (in the form of an equation) between two or more physical quantities (e.g. the differential cross section for a given process as a function of the transverse momentum etc).[5]

- *Models of data*: A model of data is the representation of a measurement outcome into a canonical form that allows – directly or indirectly – the comparison of experimental data with the hypothesis under investigation. The construction of a data model involves a variety of data analysis techniques and statistical methods, and as we shall see, it is heavily guided by background theoretical assumptions and other approximations. Depending on the hypothesis in question, a data model can take several forms such as a table, a simple numerical answer with an uncertainty estimate, or as it is most common in HEP, a function represented by a graph. Finally, although the final data model which is eventually compared with theory

---

[4] This feature will become more salient in our discussion of the various phenomenological models of dark matter in the next chapter.

[5] Note that the different nature of a model's prediction may require a different representational medium to b expressed – e.g. a graph, a mathematical equation and so on.

is typically a ‘simple entity’ as Suppes pointed out, the construction of this entity often requires a number of auxiliary intermediate data models.

- *Models of experiment*: A model of an experiment is a blanket term referring to every possible modelling activity that facilitates the completion of a measurement process in an experiment, and allows the connection between the final data model and a theoretical hypothesis. In high energy collision experiments this includes the physical models for calculating the interactions of particles with the different parts of the detector, any kind of simulation modelling that provides the basis for necessary calculations (event generators, detector simulations, pseudo-experiments with Monte Carlo simulations etc.), models of data acquisition, and finally, the various statistical models used for the analysis of data.

Measurement is understood in this context as the experimental activity which leads to the quantitative attribution of the value of a targeted physical quantity, typically represented as a theoretical parameter or variable in an idealized experimental model. Eran Tal (2017a, p.240) describes measurement in terms of two levels: the physical interaction between the target of the measurement and the measuring instrument, and the *model of measurement*, which is an abstract and idealized representation of this physical interaction. The attribution of values to various parameters from the sub-detectors of the LHCb to be described in Sections 4.3 and 4.4 based on the experimental models of the interactions between the products of the collisions and the detector, nicely illustrates Tal’s description of measurement in terms of the concrete physical interactions in the detector and the abstract models of measurement – or as we shall call them here – models of experiment. For the purposes of our discussion, it is also useful to follow Tal (2017b) and distinguish between *instrument indications* and *measurement outcomes*. The former are properties of the final states of measuring instruments after a measurement is completed such as the numerals appearing on the display of a measuring device, and are often understood as providing the raw data of the experiment. The latter are knowledge claims about the value of a physical quantity attributed to a physical process such as the claim that ‘the mass of the top quark is $M_{top} = 172.85 \pm 0.714(\text{stat.}) \pm 0.85(\text{syst.})\ GeV/c^2$’. As we shall see, it is the measurement outcomes and not the instrument indications that are represented by data models.

This slightly modified version of the hierarchy of models account, even in its crudest form, nicely captures the relationship of theories and experiments in high-energy physics. Background theory, be it the Standard Model, or any other new physics theory or model to be put to the test, such as Supersymmetry, provides its predictions via its models in the form of empirically testable hypotheses. Large scale scattering experiments are then designed and carried out based on these theories, yielding a huge volume of raw data which is eventually turned into a simple data model which is comparable to the theoretical hypothesis. The acquisition of data and the construction of the final data model is unavoidably carried out with the help of various theoretical assumptions and other intermediate experimental models. The purpose of the following sections is to illustrate how this modified account of the hierarchy of models can be applied for the description of experimental tests of the theoretical principle of Lepton Flavour Universality in HEP via the so-called B-anomalies in the rare decays of B-mesons.

4.2 B-anomalies and Lepton Flavour Universality

The Standard Model (SM) of particle physics is by far the most empirically successful physical theory for the fundamental building blocks of visible matter and the interactions between them. However, despite its tremendous empirical success, the Standard Model is also undermined by a number of experimental results that consistently deviate from its theoretical predictions.[6] One type of such results, which will be the focus of the present case study, concerns the so-called B-anomalies in the rare decays of B-mesons that are one of the main areas of study in the indirect searches for new physics at the LHCb experiment at CERN.[7]

A currently very active research programme of indirect searches with rare decays concerns the decay of quarks via a *flavour changing neutral current* (FCNC).[8] FCNC processes are only permitted by the SM at the loop-level, and thus they are significantly suppressed with respect to flavour changes mediated by charged currents, providing this way a fertile ground for indirect searches for New Physics. A prime example of an FCNC interaction that has received much attention by experimental physicists is the

---

[6] For more details on the limitations of the Standard Model and the searches for new physics at the LHC see Virdee (2016) and Rappoccio (2019).

[7] As opposed to direct searches, which aim at the observation of hypothetical new particles via their production in scattering experiments, indirect searches for new physics concern the performance of precise measurements of observables in (usually rare) scattering processes, by analysing large volumes of data on observables related to these decays.

[8] According to the standard model, the probability of a decay process whereby quarks are allowed to change flavour (e.g. processes where an up quark can decay to a down quark) is significantly higher in interactions where their electric charge is not preserved. That is, quark flavour changes via electrically charged paths mediated by the $W^{\pm}$ boson are much more likely to occur as opposed to FCNC interactions where flavour changes are mediated by the neutral Z boson.

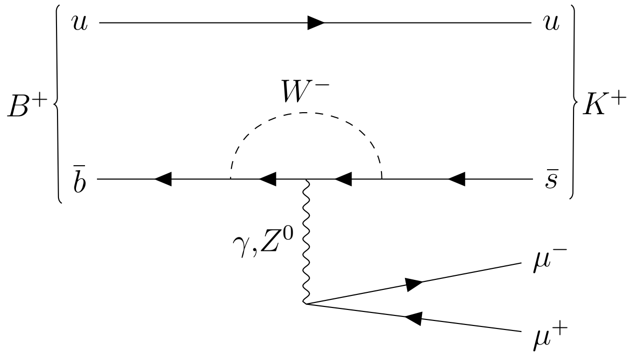

Figure 4.1: Feynman diagram of the dominant contribution to the B-meson decay $B^+ \to K^+\mu^+\mu^-$ within the Standard Model. The diagram illustrates the rare process in which a B^+ meson ($u\bar{b}$) decays into a K^+ meson ($u\bar{s}$) and a pair of muons.

semileptonic decay of b quarks into a strange quark and a pair of leptons ($b \to s\ell^+\ell^-$) which characterises rare decays of B-mesons such as the processes $B^0 \to K^{*0+-}$ and $B^+ \to K^+\mu^+\mu^-$ (Fig.4.1). This specific type of rare decays is particularly interesting for searches for New Physics since the large mass of the b quark with respect to the energy scale of QCD allows the separation of strong and electroweak contributions and, consequently, the prediction of decay rates and angular distributions with small theoretical uncertainty.

The term B-anomalies refers to a set of observed experimental results of various observables of B-decays displaying tensions with the standard model predictions at the 2-3 sigma level. The overall consistency of these results is interpreted by many physicists as a hint for the presence of new physics in these decays and hence, the accumulation of further data and the precise measurement of these observables via the appropriate data models is of ultimate importance for the development of new physics beyond the Standard Model. A particular observable in these anomalies is the R_K ratio that features in tests of *Lepton Flavour Universality* (LFU) (Bifani et al. 2018; Muller 2019). LFU is a theoretical principle of the Standard Model which stems from the fact that, apart from their mass differences, the three charged leptons (electrons, muons and taus) are identical copies of each other, and thus the electroweak coupling of the gauge bosons to leptons is independent of the lepton flavour. In practice, this means that according to the SM, electrons couple to photons, Z and $W^{\pm}$ bosons in the same way the muons and taus do. As Suppes pointed out however, this is a general theoretical principle that does not directly correspond to an experimental observable and thus a theoretical model is needed in order to convert this general theoretical principle into an empirically testable hypothesis.

The most straightforward way to do this is by constructing a model whose mathematical framework features the ratio of the branching fractions between two different B-decay processes with different flavours of leptons in their final products. The main advantage of this formulation is that, as opposed to the individual branching fractions, the various uncertainties due to the complicated QCD calculations cancel out, allowing the standard model to provide much more accurate predictions. Since the electroweak couplings of all three charged leptons are the same, the ratio of branching fractions for $B^+ \to K\mu^+\mu^-$ and $B^+ \to Ke^+e^-$ decays is naively expected to be unity, and it can indeed be calculated theoretically with high precision in a given range of the produced dilepton mass squared $q^2_{min} < q^2 < q^2_{max}$. For the particular processes, this ratio is defined by

$$R_K = \frac{\int_{q^2_{min}}^{q^2_{max}} \frac{d\Gamma[B^+ \to K^+\mu^+\mu^-]}{dq^2} dq^2}{\int_{q^2_{min}}^{q^2_{max}} \frac{d\Gamma[B^+ \to K^+e^+e^-]}{dq^2} dq^2} \tag{4.1}$$

where Γ is the q^2-dependent partial width of the decay.[9] In the low region range for the dilepton squared mass $1.1 < q^2 < 6.0$ GeV2/c^4, this ratio is predicted by the Standard Model to be unity with $\mathcal{O}(1\%)$ precision (Bordone et al. 2016). This theoretical prediction constitutes the theoretical hypothesis:

$$\mathrm{H}_0: \quad R_{K^+}[1.1, 6, 0]^{SM} = 1.00 \pm 0.01_{QED}$$

which is eventually compared to the model of data. The QED subscript indicates the origin of theoretical uncertainties due to QED effects and the numerical interval corresponds to the dilepton mass squared range.

The ultimate aim of 'the maddeningly diverse and complex experience' which constitutes the experimental test of LFU is to construct a data model of the R_K ratio: a simple entity in the form of a numerical result, subject to statistical and systematic uncertainties, which is comparable to H_0 in a precise and mathematical manner. The next section provides a brief description of the data processing system of the LHCb in order to facilitate the discussion to follow on the rather complicated process of constructing a data model for the R_K ratio. As we shall see, theory guides the observation

---

[9] The width Γ of a decay is related to a particle's lifetime τ and expresses the uncertainty in its mass m according to the Heisenberg uncertainty principle: $\Delta\tau \times \Delta m \sim \hbar/2\pi$. The faster a particle decays, the larger its decay width is and thus the uncertainty in its mass. The partial decay width expresses the uncertainty of a particle's mass with respect to a particular decay channel.

and data acquisition process in three different levels: an all-encompassing fundamental level independent of the specifics of the experiment, an intermediate level concerning the physical processes in the detector, and a more restricted level regarding the specifics of the quantities to be measured in a given experiment.[10]

4.3 Data processing at the LHCb

The LHCb experiment at CERN is currently the largest experiment in physics for the study of rare B-decays. It is specifically designed to profit from the enormous production rate of b quarks in proton-proton collisions at the Large Hadron Collider (LHC) which happen at a rate of around 3×10^{11} per fb^{-1}.[11] The LHCb detector collects about 25% of the b quarks produced in these collisions, and provides the necessary data for making precise measurements of various observables related to the rare B-decays.

The detailed study of these processes requires the determination of the various properties of the final state particles and their kinematics. In order to determine these properties and allow the full reconstruction of an interaction process, a number of different quantities need to be measured including the charge of a particle, its flavour, the momentum vector, and for short lived particles, the production and decay vertex. Since no detector can simultaneously measure these quantities, large detector systems such as the LHCb detector, are typically made of various specialised sub-detectors, each performing a different task. The various sub-detectors of the LHCb detector can be grouped into two complementary sub-systems: the Track Reconstruction system and the Particle Identification system. As the name suggests, the systems involved in track reconstruction aim in reconstructing the trajectories of charged particles in a collision event by combining information from the ‘hits’ recorded in the various sub-detectors. Once the tracks are reconstructed, the Particle Identification (PID) system derives further information from its sub-detectors in order to associate the tracks with a specific particle species. Together with the momentum information provided by the tracking system, the PID also allows the energy of a charged track to be computed using the relativistic energy-momentum relation $E^2 = p^2c^2 + m^2c^4$.[12]

---

[10] The content of this section was derived from Teubert (2016), Blake, Lanfranchi & Straub (2017), Cabdevilla et al. (2018), Lionetto (2018), Mauri (2018), Lisovskyi (2019) and Humair (2019).

[11] One inverse femtobarn (fb^{-1}) corresponds to approximately 100 trillion ($\sim 10^{14}$) proton-proton collisions.

[12] For a detailed description of the LHCb detector see the official publication from the LHCb collaboration (LHCb collaboration, 2008).

4.3.1 The three levels of theory-ladenness

Already one may notice here the first and most general level of the theory-ladenness of observation. The overall design and operation of the track reconstruction and particle identification systems at the LHCb (as well as of any other large scale experiment) is based on a number of physical principles that are considered to be fundamental and are expected to hold in any possible new physics theory to be constructed based on these data. These general principles enter the observational process in the form of various implicit and explicit assumptions which lie at the core of almost every experiment in physics and concern the most fundamental facts we know about nature, such as the conservation of energy and Einstein's mass-energy equivalence principle. This type of theory-ladenness is universal across a particular field of physics and is independent of the aims and quirks of any particular experiment.

The second level of theory-ladenness of observation in collision experiments concerns the physical processes behind the production of 'hits' in the detectors and the identification of particles. During a proton-proton collision event, hits are produced in the various trackers by the energy loss of the traversing particles due to their interaction with matter. The two main physical processes that occur in the detectors are inelastic collisions of the products with the atomic electrons and elastic collisions with the nuclei of the atoms of the detectors' material, leading to the phenomena of ionisation and multiple Coulomb scattering respectively. Theory-ladenness appears at this stage by offering the various physical models for calculating the effects of these physical processes on the detector.

As aptly noted by Beauchemin (2017, p.299), quite often there are more than one competing models about the nuclear interactions between charged hadrons and the material of the detector. However, these competing models, although empirically equivalent, affect the simulation of the detector and the selection of data in different ways, giving rise to different results. In other words, the extrapolation and interpretation of data, and consequently the form of the final experimental result as a data model, depends on the choice of the model for the underlying physical processes in the detector. This fact raises the worry of a possible vicious circularity due to the theory ladenness of data selection. If the result depends on the arbitrary choice between several empirically equivalent models, what validates the objectivity of a given result based on a particular model? As will be shown in Section 4.4.4, the solution to this problem is achieved by separately calculating the effects of each model to the measurement and including them in the systematic uncertainties of the final result.

The third and most specific level of theory ladenness concerns the theoretical principles and assumptions that are specific to the aims of the particular experiment which will be described in the following section. These assumptions basically determine (i) the selection criteria for distinguishing the data from what are considered to be the 'interesting events', i.e. events related to the two decays consisting the R_K ratio and (ii) the vast majority of theoretical and mathematical calculations involved in the derivation of the final result. The suggested tripartite distinction of theory-ladenness presented here partially overlaps with Karaca's (2013) two-fold distinction between the strong and the weak sense of theory-ladenness of experimentation, albeit with an additional intermediate layer. Karaca describes the strong sense of theory ladenness experimentation as the continuous guidance of an experiment by some theoretical account with the aim of ascertaining the conclusions of the same account. This strong sense of theory-ladenness is captured by what we have labelled here as the third and most specific level of theory ladenness which essentially determines the collection and further refinement of data at the LHCb trigger system, in order to construct an appropriate data model to be compared with the theoretical hypothesis in question. The weak sense is described by Karaca in a broader context, as the utilization of theoretical considerations that have no guiding power on the progress of the experimental process.

4.3.2 The LHCb trigger system

Before moving to the analysis of the data modelling process for the R_K ratio it is useful to give a brief description of some technical details regarding the data processing system of the LHCb. The rate of visible collisions at the LHC, i.e. the number of recorded events per second, is currently between 10 and 20 millions ($\sim$ 13 MHz in Run II).[13] This number is simply too big to allow every single event to be stored for further analysis and thus, a filtering system is required to select the interesting events by filtering out the events containing various well-studied physical processes that are unrelated to the specific aims of the experiment. For the LHCb experiment, this amounts to the selection of the events that are most likely to contain a B-meson or a D-meson, since, in addition to the study of the rare B-decays, LHCb is also dedicated to the study of D-decays (decays of heavy D-mesons consisting of at least one charm quark/antiquark) and CP violations. The selection of these events is completed in two levels by the LHCb Trigger system, and is

---

[13] In the jargon of particle physics, the recording of an *event* amounts to the recording of all the products from a given collision. Run I and Run II refer to the different periods of operational running for the LHC under different conditions. Run I took place in 2009-2013 and Run II in 2015-2018. Run III is scheduled to take place in the years 2021-2023.

based on the information from the various subsystems of the detector.

Practically speaking, the ultimate task of the software algorithms connected to the tracking and the particle identification system of the detector is to attach values to several variables related to the kinematics of the interactions (momentum, energy, mass etc), their topology (scattering angles, flight distance, impact parameter) and the nature of the particles. Two simple examples of such variables are the binary `isMuon` variable which depends on the number of hits in the muon stations associated with a track, and the $DLL_x(t)$ variable which corresponds to the likelihood of a track t to belong to a particle species x rather than a pion. These variables can be produced based on information from either a single subdetector or by combining information from several detectors. The job of the trigger system is then to take these variables as inputs and, based on a number of selection criteria that are also known as *cuts*, decide whether a given event is of interest or not.

In order to be able to distinguish the interesting events from the various processes taking place in the LHC, the trigger system of the LHCb is programmed to search for the characteristic signatures of hadrons containing b or c (anti)quarks, which give rise to the heavy flavour decays in which we are interested. The three most significant signatures of these hadrons are (i) their large lifetime, which results in long flight distances compared to the resolution of the detector, (ii) their large mass, which results in high transverse momentum P_T of the product particles,[14] and (iii) the existence of muons in the final state of several key decay modes of these hadrons, such as the $B^+ \rightarrow K^{++-}$ decay in which we are interested in for the measurement of the R_K ratio (Head 2014). This is where the third – aim specific – level of theory-ladenness becomes apparent: the specification of these signatures for the data selection process is largely driven by various theoretical assumptions for the nature of these decays based on the existing background theoretical knowledge. This fact is also related to what was said earlier in Section 4.1 about the necessary connection of data acquisition with a clear research question. The fact that the ultimate purpose of this particular experiment is to test LFU via the R_K ratio specifies which events are of interest for this purpose, and eventually determines the choice of the most appropriate selection criteria to distinguish these events.

The first level of the LHCb trigger system is completed by the Low Level trigger (L0). L0 is a hardware based trigger and its task is to reduce the data output from ~ 13 MHz to approximately 1 MHz at which the LHCb detector can be read out. Contrary

---

[14] Transverse momentum is the component of momentum transverse (i.e. perpendicular) to the beam line.

to what one might expect, the selection criteria at this first level are not purely theory-laden, rather they are mainly determined by a number of technical limitations. In order to achieve the goal of 1 MHz, the L0 trigger needs to take a decision for every event in a very short amount of time (4μs), and for this reason, it only receives information from the muon system and the calorimeters, as these are the only sub-detectors able to provide information in such a short amount of time. Once this information is received, the trigger algorithm discards all events with too many hits in the SPD detector since such high occupancy events would require an excessive fraction of the available processing time at the next level of the data process. After these criteria are applied, L0 proceeds to a coarse-graining of the interesting events by selecting muons with a high transverse momentum p_T and other events with high energy deposits in the calorimeter. The thresholds for these cuts are not fixed, rather they are constantly changing according to the data-taking conditions of the experiment, even during the same year or Run.

The second level of the trigger system is completed in two stages by the High Level triggers (HLT1 and HLT2) of the detector. These are software based algorithms and their task is to further reduce the amount of data in order to be stored onto servers at the CERN Data Centre and distributed to physicists for analysis. In Run II, HLT is programmed to reduce the data rate from the 1MHz output of the L0 trigger to 12,5 kHz which is low enough to be permanently stored on disks. During the first stage of the High Level triggering, HLT1 receives information from the tracking system and proceeds to a partial event reconstruction by applying various selection criteria based on the impact parameter of the events, the quality of the tracks and the transverse momentum.[15] This process reduces the data output to $\sim$70 kHz and passes the selected tracks to HLT2. The HLT2 algorithm then performs a full reconstruction of all the selected events that satisfy $P_T > 0.3$ GeV independently of their impact parameter or matching hits in the muon chambers. The overall process of reducing the amount of data from the LHC collisions to a manageable dataset to be distributed widely is illustrated schematically in Fig.4.2.

This pragmatic dimension of the data acquisition process is nicely captured by Bokulich and Parker in their discussion of the *problem space* in data modelling, in which the goal is to achieve a particular purpose of interest guiding the construction of the model (2021, p.12). As Bokulich and Parker note, the final properties of the data model are jointly determined by the different dimensions of the problem space, namely the representational relationship between the data and the target of the experiment, the data

---

[15] The impact parameter of a particle, typically denoted by χ^2, is related to the angle of scattering, i.e. the angle at which a particle is deflected by a another particle after collision, and is used in particle identification to tag flavours to the particles.

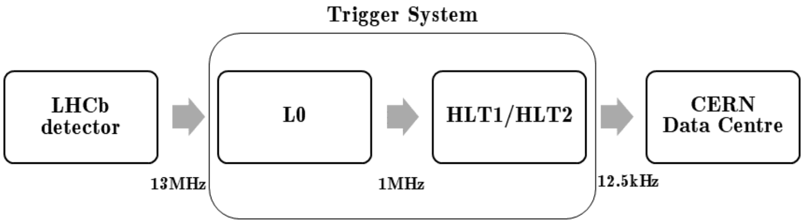

Figure 4.2: Schematic representation of the various stages of the data acquisition process at the LHCb. The numbers below the arrows indicate the size of data before and after each level.

users, the adopted methodology and the background circumstances of the experiment. Given that the purpose of our case study is to test LFU via the R_K ratio, the final form of the data model is indeed jointly determined by a number of theoretical and pragmatic factors including the choice of models for the interactions of product particles with the detector, the available computational time and power, the storage capacity at the CERN Data Centre, the reconstruction of events so that they are amenable to statistical analysis and so on.

The completion of the High Level triggering process marks a significant milestone where the vast majority of the available data from the proton-proton collisions at the LHC is discarded *irretrievably*, mainly due to the technical limitations of the data processing system both in terms of the data-processing time and the store capacity of CERN's Data Centre. The reduction of data from 13 MHz to the final 12,5 kHz that eventually becomes available to the users means that about 99,9% of the available data from visible collisions never reaches the physicists' desks for further analysis. Add this to the fact that only about 1% of the actual collisions in the LHC provide products that end up in the detectors, and it is not hard to see that the otherwise huge amount of data that eventually gets stored and analysed by physicists is only a minute fraction of the potentially available information provided by the proton-proton collisions at the LHC. Even though extreme care is taken to make sure that the data collected correspond to the events containing new physics, it is widely acknowledged by the physics community that a large amount of information containing hints to new physics is permanently lost during this process.

This brings us to the final aim of the overall experimental process which is the acquisition and organisation of data for the construction of the data model representing the results of the experiment. Part of this process concerns the determination of the selection criteria for the collection of data, whereas another part takes place only once a sufficient amount of data for the study of a particular phenomenon becomes available at the CERN Data Centre. The datasets are then widely distributed to the scientific community on an international scale by the Worldwide LHC Computing Grid (WLCG) for further statistical and mathematical analysis. Given the huge amount of data required to produce reliable results, it should be stressed that the overall process of deriving an experimental result from the available data is typically a *non-linear* and laborious activity of constant refinement and revision, which usually takes years of collaborative work to complete. Nevertheless, it can be characterised by four main stages which will be the focus of the next section. Although in practice these stages do not follow a clear chronological order and are not always clearly distinguishable, they nicely capture the most essential procedures for the construction of a data model in HEP.

4.4 Constructing data models for the R_K ratio

This section provides a description of the four main stages for the construction of the data model of the R_K ratio as it was recently presented by the LHCb collaboration (2019; 2021): (i) data selection (ii) efficiency calculations (iii) data fitting and (iv) uncertainty calculations. The analysis of these four stages illustrates the importance of considerations for the construction of a data model with respect to the data acquisition criteria, the complicated calculations of the performance and efficiency of the detector with the help of simulation, the fitting of finite data to continuous functions via statistical analysis, and the evaluation of possible errors during the measurement process. As we shall see, each stage of this procedure is, in its own way, replete with various underlying theoretical assumptions, giving further credence to the idea that observation in HEP is highly theory-laden.

Before we proceed with the description of these four stages, the first thing to note is that the measurement of the R_K ratio is conducted in the low region of the produced dilepton squared mass ($1.1 < q^2 < 6.0$ GeV2/c^4) since this is considered to be the 'cleanest' region for the study of the rare semileptonic transitions $b \to s\ell^+\ell^-$ characterising the B-decays. B-decays such as the one shown in Fig.4.1 are not only mediated by the semileptonic $b \to s\ell^+\ell^-$ transition, but can also proceed by various other processes known as intermediate resonances. The chosen region is considered to be the cleanest in the sense that it is mainly dominated by the rare B-decays constituting the ratio, as opposed to other regions of q^2 that are mainly dominated by resonant decays, making the distinction of signal and background even harder. This is another example of the third and strongest sense of theory-ladenness, whereby theory dictates the scope of the observation by determining the most appropriate region for the study of B-decays.

As already noted in Section 4.2 (Eq.4.1), the R_K ratio is defined as the ratio between the branching fractions of the decays $B^+ \to K^+\mu^+\mu^-$ and $B^+ \to K^+e^+e^-$, and thus the measurement of the ratio *prima facie* requires the measurement of the two branching fractions. However, a first challenge arises due to the large amount of bremsstrahlung emission by the electrons in the detector which significantly complicates the reconstruction of electron tracks compared to the muons. The way to deal with this imbalance in the reconstruction and identification of electron and muon tracks is to instead calculate the R_K ratio as a double ratio of the branching fractions:

$$R_K = \frac{\mathcal{B}(B^+ \to K^+\mu^+\mu^-)}{\mathcal{B}(B^+ \to J/\psi(\to \mu + \mu^-)K^+)} \bigg/ \frac{\mathcal{B}(B^+ \to K^+e^+e^-)}{\mathcal{B}(B^+ \to J/\psi(\to e^+e^-)K^+} \tag{4.2}$$

The $B^+ \to J/\psi(\to \ell^+\ell^-)K^+$ decays are intermediate resonant decays in which a B-meson decays into a K^+ meson and a J/ψ meson with a very short lifetime which further decays into a pair of muons or electrons providing the same products as the nonresonant B-decays constituting the R_K ratio. The use of this double ratio exploits the fact that the resonant $B^+ \to J/\psi(\to \ell^+\ell^-)K^+$ decays have already been observed to have lepton-universal branching fractions with great precision, and thus they do not affect the true value of R_K. As opposed to Eq.(4.1), Eq.(4.2) requires the detection efficiency for the nonresonant decays to be known only relative to the corresponding resonant decay and thus any systematic uncertainties related to bremsstrahlung emission are significantly reduced. The upshot is that the successful measurement of the R_K ratio, not only depends on a large number of underlying theoretical assumptions, but also on the assumption that any previous experimental results that are needed for the measurement are solid and reliable.

4.4.1 Selection criteria

The first stage in the construction of the R_K data model concerns the determination of the selection criteria to be applied to the trigger system in order to distinguish the signal – i.e. the events of interest that contribute to the ratio – from the background – i.e. the unrelated events in the collision with similar signatures. The involved strategies during this stage are determined at each step according to the source and the specific characteristics of each type of background based on the existing theoretical knowledge. For instance, a particularly invasive form of background comes from the mis-identification of pions as leptons in the $B^+ \to K^+\pi^+\pi^-$ decays, which are 30 times more frequent than the B-decays constituting the ratio. The suppression of this background is achieved by applying a combination of cuts for the `isMuon`, the DLL_μ, and the DLL_e variables in the particle identification algorithms. In general, the choice of these cuts is based on a combination of both theoretical and pragmatic criteria regarding the expected behaviour of the detector with respect to each type of background. The underlying assumption is that there is sufficient knowledge of the nature of different types of the background processes which produce signals that could potentially be mis-identified by the detector as coming from the rare B-decays.

The main challenge at this stage, is what Franklin (1998, 2015) calls ‘the problem of cuts’, which stems from the possibility that the experimental result simply reflects the choices of the particular cuts on the triggering system. In other words, the worry is that certain combinations of cuts will give rise to different sets of results and there is simply no way of knowing which of these combinations provides a genuine unbiased result. The situation becomes worse in cases where the effects of the cuts to the result are known to the experimenter in advance, and hence, the idea of producing a desired outcome may distort the objectivity of the experimental results. As Franklin notes, the experimenter’s bias is mitigated by applying the practice of ‘blind analysis’, in which the experimenters analysing the data do not know the result until the analysis method is finalized, following an extended peer review within the collaboration. Beacheumin (2017) adds that the solution to these problems also comes from the implementation of systematic uncertainties in the result, which will be further discussed in Section 4.4.4.

Once the first stage of calculating and applying the selection criteria for distinguishing the relevant decays for the R_K ratio is completed, the measurement of the ratio requires the calculation of two types of quantities: the efficiencies, ε, for selecting each one of the decays and the yield, N, of each decay mode, which is the number of recorded events contributing to the ratio. The calculation of these two types of quantities constitutes the second and third stage respectively for the construction of the R_K data model.

4.4.2 Efficiency calculations

The second stage of the data modelling process concerns the calculation of the detector’s efficiency during the triggering, reconstruction, and identification processes. These efficiencies are usually integrated in the total efficiency of the detector, ε_{tot}, which can be defined as the fraction of the events registered and correctly identified at the detector, with respect to the actual number of events produced by the proton-proton collision in the LHC. The knowledge of these efficiencies is essential, since in order to know the true value of the ratio between the two yields, it is clear that we must first be in a position to know how many of the rare B-decays that actually occur are eventually recorded by the detector and become available for analysis. The calculation of the ‘true number’ of rare B-decays is a crucial yet challenging aspect at this stage. Given that the only way to detect and count these decays is via the – imperfect – detectors, how is it possible to know how many of these decays are eventually recorded? The answer is via simulation.

The overall process of calculating the efficiency of the detector by simulation can be described in three steps. The first step is to provide a complete list of all the particles that come out of a certain physical process, including the ones that are stable enough to interact with the detector. This is made possible by various software algorithms that are known as *event generators*. When combined, event generators provide a complete description of all the particles that come out from a collision between protons in which a B-meson is produced, providing this way the necessary knowledge for the expected yield of rare B-decays.

Once the events are generated, the next step in this stage is to simulate the path of the produced particles in the various parts of the detector, in order to model the detector’s response. This process requires the construction of a detailed digital map of the LHCb detector in a language that is readable to the software. Ideally, this map would include every single wire and pipe of the detector ensuring that the simulation provides accurate results, however, this would require an unrealistic amount of processing time, and thus various approximations are used. This part of the simulation also involves the implementation of various physics models in the software, describing the different physical processes that are expected to take place in the detector (bremsstrahlung, ionization, multiple scattering etc.) according to the background theory. Once again, it should be noted that it is not possible to include in the simulation every single physical process that is expected to take place (this would require the simulation to run for a tremendous amount of time) and thus, the physical models are chosen on a pragmatic basis, taking into account limitations on time and computational power. The final output of this second step in the simulation is a large database with information about energy deposits in the detector including their times and locations.

The third and final step in simulation is the *digitization* of data. This is the process whereby the available data from simulation is converted in the same format as the data provided by the experiment electronics and the detector’s data acquisition system. The idea is to convert the information about the energy deposits from the simulation into whatever it is that the detector actually reads – i.e. voltages, currents and times.

Moreover, this is the stage where various other interesting detector effects are also taken into consideration with the help of various models, such as the difference in light collected from a scintillator tile in the calorimeter depending on whether the energy is deposited in the middle or in the edge of the tile. The final result is a simulated dataset that has the exact same format as the data coming out of the detector's data acquisition system, and for which, as opposed to the real data, there is precise knowledge of the physical processes that generate them. This allows the calculation of the efficiencies of the tracking and particle identification systems of the detector. After digitization, the simulated data follow the exact same path through the trigger system just as the real data, allowing this way the calculation of the efficiency of the trigger system as well.

This procedure is not immune to problems either however. Even though the simulation is considered to provide a good estimate of the detector's efficiency in real data acquisition, it is still possible to have discrepancies between the simulation-calculated efficiency and the true efficiency of the detector. This may happen for instance due to technical problems during data acquisition from real collisions that are not taken into account in the simulation, or poor modelling of certain aspects of the detector in the simulation software (for instance, it is known that the performance of the RICH detectors and calorimeters is not accurately simulated by the LHCb software). These discrepancies are often corrected by a data-driven method called 'tag & probe' whereby the simulation efficiency is revised based on data calibration samples from other well-studied decays.[16]

Another way to test the validity of the methods of determining the efficiencies is to take advantage of the (experimental) fact that the nonresonant J/ψ decays are known to respect LFU at a good level of precision. The measurement of the single J/ψ ratio by the very same process of measuring the double ratio is indeed considered as a solid cross-check for the robustness of the efficiency calculation method, since it does not rely on the double ratio's cancellation of systematic uncertainties. The underlying assumption here is that the previous results of LFU in the single resonant decays do not suffer from any sort of unknown uncertainties. Known uncertainties on the other hand, are, as expected, integrated and reflected in the systematic uncertainties of the new results via a process known in statistics as 'error propagation'.

The various models described during this stage – i.e. detector layout models, models of the physical processes and other effects in the detector, models of data flow in the detector etc. – are all part of the class of experimental models described in Section 4.1, whose task is to facilitate the connection between the theoretical and the final data

---

[16] For a detailed description of the 'tag & probe' method see Archilli et al. (2013).

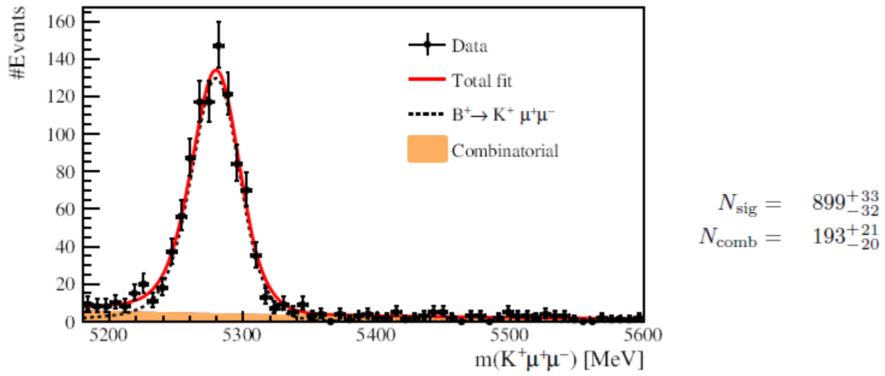

Figure 4.3: Fit to $m(K^+\ell^+\ell^-)$ for $B^+ \to K^+\ell^+\ell^-$ events in the Run II data, along with the contribution from combinatorial background. The extracted values for the signal yield N_{sig} and the background N_{comb} are displayed at the right of the figure (Humair 2019).

model of the experiment.

4.4.3 Data fits

As already mentioned, in addition to the detector efficiencies, the measurement of the R_K ratio requires the calculation of the yields N of the decays of the ratio. This is achieved in the third stage of data modelling via the process of data fitting. In general, data fitting is the mathematical process of finding a function that *best fits* a number of data points (i.e. the process of 'fitting the curve'), with the aim of determining or estimating the values of various unknown parameters affecting the collection of data. As noted by Suppes (1962, p.253), one of the most profound complications in the reconciliation of data and theoretical predictions is that the former are of a discrete and finite nature, whereas the latter are typically continuous functions or infinite sequences. Data fitting is the mathematical tool for solving this tension by finding the most appropriate (continuous) function that best describes the finite sequence of data collected in the experiment.

In the context of the measurement of the R_K ratio, the fits are performed to the data for the combined mass $m(K^+\ell^+\ell^-)$ in each decay, providing this way a probability distribution for the mass of the B-meson. This distribution is considered to be the best description of the set of observations x_i, given that these observations are also affected by the presence of residual background (i.e. background that evades the data selection process). Once the fit is performed, the probability density function is re-parametrised so that it is a function of the relevant yield N and R_K, and maximum likelihood estimations are then performed to find *the values of the yields for the signal* N_s *and background* N_b *for* *which it is most likely to observe the given masses* $m(K^+\ell^+\ell^-)$ in each decay process. For instance, Fig.4.3 illustrates the fit performed to the $m(K^+\mu^+\mu^-)$ data in order to extract the yield of the nonresonant decay $B^+ \to K^+\mu^+\mu^-$. This is an example of an auxiliary data model needed for the construction of the final data model representing the R_K ratio. The extraction of the yield from these fits involves the use of specialized software algorithms both for the determination of the shape of the curve and the maximum likelihood estimation of the parameters, taking into account all possible contaminations to the fit from background contributions.

Maximum likelihood estimation (MLE) is one of the most popular statistical methods for calculating unknown parameters such as the yields of decay processes in high energy physics experiments. Roughly speaking, given a probability density function $f(x_i;\theta_i)$ describing a set of observations x_i, that are characterised by a set of parameters θ_i, MLE is a method of finding the values of θ_i that make the data most likely. What this means in practice, is that the final value of R_K which is eventually compared to the theoretical hypothesis H_0, *is itself* a hypothesis as well, which is, nonetheless, derived from the available experimental data on the basis of various mathematical criteria. What makes MLE a popular method in HEP is the fact that compared to other estimation methods, it is characterized by a number of 'good' statistical properties such as consistency, small bias and robustness.[17]

For completeness, let us note here that the most likely value of the R_K ratio given the available data from the most recent measurement (LHCb collaboration, 2021) was found to be

$$R_K = 0.846^{+0.042+0.039}_{-0.013-0.012} \tag{4.3}$$

where the first uncertainty is statistical and the second systematic. The fourth and last part in the construction of data models concerns the determination of these uncertainties, which, as we shall see, are a very important and indispensable part of a data model.

4.4.4 Uncertainty calculations

The attribution of statistical and systematic uncertainties in a HEP experimental result can be understood as a way of quantifying possible errors in the data taking process.

---

[17] In short, an estimator of a parameter is said to be *consistent* if it converges, in probability, to the true value of the unknown parameter as the number of measurements tends to infinity. The *bias* of an estimator is the average deviation of the estimate from the true value over an infinitely large number of repeated experiments. *Robustness* is the property of an estimator to have limited sensitivity to the presence of outliers in the data. The full mathematical definitions of these properties can be found in Lista (2016, Ch.5).

This understanding reflects the seemingly more popular ‘error approach’ in HEP, whose objective is to determine an estimate of a quantity which is as close as possible to the unique true value of the quantity. This is opposed to the ‘uncertainty approach’ whose objective is to determine an interval of values which can be equally assigned to a quantity with relatively high confidence, and can be understood as a way of quantifying doubt during a measurement process.[18] In the case of the R_K ratio, the preference to the error approach is reflected by the expression of the result as a single numerical value – which is supposedly as close as possible to the real value of the ratio – associated with statistical and systematic uncertainties.

Generally speaking, in a HEP experiment, there are six main sources of uncertainty: (i) the intrinsic probabilistic nature of the underlying quantum field theory, (ii) the theoretical uncertainties involved in the calculation of various quantities due to highly complicated (usually QCD related) theoretical calculations (iii) the various measurement errors that are present in the data taking process even without taking into account any quantum effects, (iv) the variability in the selection of different models and different measurement methods in the experiment (v) the experimenter’s insufficient knowledge about various aspects of the experiment due to limitations of cost, computational time, computational power and so on, and (vi) the simple fact that a repeated measurement may yield different results for the same quantity.[19]

These and other possible sources of error give rise to two different types of uncertainty that are typically accompanying a HEP result in the form of the data model: statistical uncertainties and systematic uncertainties. A possible way to distinguish between these two types of uncertainty on the semantic level, is to understand the former as expressing the possible fluctuations in a measurement result even when all input quantities and other factors affecting the measurement are perfectly known and stable. This means that the presence of statistical uncertainty can be attributed to the probabilistic nature of quantum field theory and other purely statistical factors, and thus its minimisation is quite often merely a matter of collecting additional data in future runs. Systematic uncertainties on the other hand, can be seen as resulting from our imperfect knowledge on various aspects of the experiment, the mis-modeling of detectors in the simulations, and the possible defects and biases of measuring instruments during the data taking procedures. A large part of the data analysis process therefore concerns the precise cal-

---

[18] See Mari and Giordani (2014) for an illuminating discussion of the error approach and the uncertainty approach in science.

[19] As aptly noted by an anonymous referee during the reviewing process of this chapter as a journal article, this list – or any list – is, of course, non exhaustive.

culation and mitigation of systematic uncertainties to the extent allowed by the available funds and the available time, which in turn will provide extra security and robustness to the final result.

In a recent study on uncertainties in HEP, Staley (2020) aptly notes that the distinction between statistical and systematic uncertainties in HEP is rather opaque.[20] The main reason for this ambiguity stems from the fact that the sources of systematic uncertainty in a measurement are often unknown and difficult to distinguish from statistical uncertainties. Moreover, they often require a different method of evaluation, which in turn makes the combination of systematic and statistical uncertainties in the final result problematic. In order to resolve this lack of consensus, particle physicists have developed an extensive literature on the treatment of systematic uncertainties providing possible definitions and practical guidance on methods of statistical evaluation.[21] Barlow (2002) for instance, provides two conflicting definitions of systematic error by 'widely read and accepted authors' and shows how different measurements in HEP reflect these two definitions. He then concludes his paper with a set of practical advice for practitioners.

Given this ambiguity, in practice, the lack of consensus on the distinction between statistical and systematic uncertainties is usually resolved by simply stating the sources of statistical and systematic uncertainties in a published result.[22] In the case of the R_K ratio, '*by convention*, the uncertainty on R_K arising from the statistical fluctuations affecting [the ratio of the yields] $\frac{N_{K\mu\mu}}{N_{Kee}}$ is referred to as statistical uncertainty' (Humair 2019, p.133, emphasis added). All other sources of uncertainty are integrated as systematic uncertainties and are listed below (*ibid.*, p.138):

1. Calibration samples size
2. Kinematic reweighting
3. PID calibration
4. Trigger calibration

---

[20] The following remarks from experimental physicist Pekka Sinervo confirm this: 'the definition of these two sources of uncertainty in a measurement is in practice not clearly defined, which leads to confusion and in some cases incorrect inferences. [...] The definition of such uncertainties is often ad hoc in a given measurement, and there are few broadly-accepted techniques to incorporate them into the process of statistical inference' (2003, p.122).

[21] See for example Barlow (2002), Sinervo (2003), Lyons (2006), Wanke (2016), Bailey (2017) and references therein. Staley (2020) offers a very illustrative philosophical analysis of the various aspects of this debate.

[22] This, of course does not solve the problem of how one should evaluate and combine these two types of uncertainty as noted by Staley (2020).

5. Occupancy proxy
6. Tracking efficiency
7. q^2 and mass resolution
8. Decay model
9. Fit shape

Some of these uncertainties are related to the finite nature of the data samples while others come from various limitations in the detector, and the presence of physical effects like bremsstrahlung which significantly complicates the identification of electron tracks. The calculation of each type of uncertainty follows a different methodology according to the nature of the source, but the main idea remains the same. As we have seen, the overall data taking process for the extraction of the R_K ratio involves the utilisation of various auxiliary models and other assumptions that are necessary for carrying out the calculations leading to the final result. However, given that there is often no theoretical or empirical justification for (i) the use of one experimental model over another, or (ii) the assignment of a particular value in a parameter of a model or an assumption (e.g. a specific threshold value in the triggering system) the result is extracted several times either by varying the auxiliary experimental model or the value of a parameter within the selected model (Staley 2020, p.102). The variance in the result due to the use of different models and different parameters is then recorded as a systematic uncertainty.

Uncertainties are a crucial and indispensable part for the reliability of an experimental result but a further discussion of their nature and exact role requires a much deeper analysis which is beyond the scope of this chapter. As a closing remark, let us simply note that in addition to being a quantifiable measure of comparison between different results from different experiments, uncertainties are also a solid way of determining the accuracy and precision of a specific result. This point has been nicely illustrated by Beauchemin (2017) who emphasizes the critical role of uncertainties in determining the robustness and the validity of measurements. A measurement is robust insofar as the systematic uncertainties on the final results are ‘sufficiently small’ regardless of the source of these uncertainties. Sufficiently small is to be understood here as being significantly smaller than the order of magnitude of the physical effect to be measured. How much smaller is significantly smaller is not written in stone, however, the main idea is that the smaller the uncertainty, the more robust the result will be.[23]

Beauchemin also notes, rather interestingly, that in cases of small uncertainties, the allegedly vicious circularity of theory-ladenness in observation is not problematic, precisely because ‘its impact on the physics conclusions will be small and fully accounted for’ (*ibid.*, p.303). Beauchemin’s remarks have been further elaborated by Ritson and Staley (2020) who nicely illustrate how the identification of the assumptions on which a result depends and the further quantification of the dependence of this result on the various assumptions in terms of uncertainty calculations, jointly control the possibility of a vicious circularity at the practical level. The determination of the dependence of the result on the various theoretical assumptions in terms of uncertainty serves in discriminating amongst those model assumptions that have the highest impact on the uncertainty of the result and those whose variation introduces negligible changes. The clear separation between the statistical and systematic uncertainties, and the identification of the different sources of uncertainties in the published result as presented in the above list, nicely demonstrates how Ritson and Staley’s observations can be applied at the example of the R_K ratio.

4.5 Two dubious distinctions

Now that we have seen how the available data are treated in different ways during the various stages of the construction of a data model, we are in a position to make some remarks about the two distinctions between (i) raw and processed data and (ii) simulated and real (or signal) data. Although in both cases, the two extremes in these distinctions can be clearly defined, the transition between the two types of data in each case is, as we have seen, quite blurry. Regarding the first distinction, raw data are often defined as objects that are directly perceived by our experience without any mediating processing or influence by theory (*cf.* Harris, 2003, p.1511). If this definition is taken seriously, then it is not clear at all what should be counted as raw data in a large-scale HEP experiment. In practice, physicists tend to refer to the electronic signals produced by the physical processes in the various parts of the detector as the ‘raw data’ given to us by the proton collisions, whereas the output of the triggering system that eventually gets stored in the data centre and reaches the hands of researchers is referred to as ‘reconstructed events’.

---

[23] In the next and final chapter, we shall be concerned with the robustness of experimental results in more detail, within the context of dark matter research.

However, none of the signals produced in the detectors is actually directly perceived by the researchers at CERN. Before reaching the hands of physicists, the data from the electric signals produced at the early stages of the experiment at the Track Reconstruction and Part Identification systems described in Section 4.3, undergo a long process of refinement and reconstruction by the computer algorithms of the LHCb detectors and triggering systems. Hence, the 'first points of contact' – i.e. the reconstructed events – are long lists of numerical data about energy deposits on detectors, momenta etc, but as we have seen, these data are far from being unmanipulated and clear from any theoretical influence. The very nature of particle physics therefore makes it impossible to talk about raw data in this field in a strong sense.

This point also illustrates that the basic definition of data models as 'a corrected, rectified, regimented, and in many instances idealized version of the data we gain from immediate observation' given by Frigg and Hartmann (2016) does not really apply in the case of HEP. Nevertheless, the data in the reconstructed events are, in a sense, also 'raw', since they still need to undergo a long process of further analysis by scientists in order to reach their final form as a data model which is comparable to the theory. A more appropriate way to describe this situation is thus to say that data follow a long 'ripening' journey which starts from their birth as electric signals in the heart of detectors, and goes all the way up to the final polished form of a data model, without really worrying at which stage the data should be considered to be raw.[24] It is precisely for this reason that the novel definition of data models in Section 4.1, does not depend on a clear-cut distinction between raw and processed data and thus avoids the relevant objection discussed by Leonelli.

Regarding the folk view that sees the tampering of data as an act of decreasing their epistemic reliability, it should be obvious from our discussion so far that this does not apply to HEP experiments. Generally speaking, a dataset is epistemically reliable if the information it provides for the physical phenomenon it represents is correct. In the context of LFU tests, to say that the processing of data decreases their epistemic reliability is therefore to say that the processed datasets provide less accurate information about the possible violation of LFU in B-decays compared to their less processed counterparts. This is not true however. The successful completion of a large-scale experiment in HEP and the extraction of meaningful and reliable conclusions about the empirical adequacy

---

[24] Bokulich's comment on the blurriness of this distinction is characteristic: 'I will not engage the difficult question here of where exactly to draw the line between (raw) data and a data model. It may very well be that the distinction is one of degree with vague boundaries, rather than a difference of kind; [...] and where the line is drawn may further be context dependent' (2018, fn.25).

of various theoretical claims, *necessarily* requires the processing of data by statistical methods and computer simulations. For instance, as we have seen, the calculation of detector efficiencies (Sec. 4.4.2) involves the introduction of simulated data in the datasets which can be seen as a form of tampering the initial data. This step however, is taken to ensure that the calculated number of yields in the data fitting stage reflects the actual number of B-decays occurring in the collider and not the number of yields detected by the LHCb. Hence, the processing of data in some cases increases the reliability of the datasets in that it mitigates the impact of possible errors in the less processed datasets due to poor detector performance, computational limitations and so on.

Moreover, it is safe to say that the so-called raw data from these experiments, are not just epistemically less reliable than the processed data in some cases, but when it comes to their comparison with theoretical predictions, they are also practically useless in their pure form. The successful comparison of a theoretical hypothesis with data necessarily requires that the raw data extracted from the detectors are moulded into an appropriate form that makes them comparable to theoretical predictions in order to serve the purpose for which they are extracted. However, the raw data extracted from the first level of the triggering system are far from fulfilling this requirement. Hence, the seemingly counterintuitive claim that processed data are epistemically more reliable and more useful than the raw data obtained by experiments is actually a platitude when it comes to HEP.

As for the distinction between simulated and real data in HEP, this has been already discussed in detail by Margaret Morrison (2015, Ch.8). Morrison uses the example of the Higgs discovery to emphasize the absolute necessity of simulation, not only in calculating the efficiency of a detector, but also in almost every other aspect of the LHC experiments. Her main conclusion is that given that simulation and signal data are essentially combined during the data analysis process, the sharp distinction between simulation and experiment is practically meaningless, and that ‘simulation is as much part of the experiment as the signal data’ (*ibid.*, p.289). Parker (2017) reaches a similar conclusion arguing that the results of computer simulations that are often embedded in measurement practices can be understood as measurement outcomes of equal epistemic importance to the outcomes of real measurements.

The calculation of the efficiencies via simulation provided in Section 4.4.2 is a clear example of such cases, where simulation results are actually embedded in real measurement outcomes in a way that makes it practically impossible to distinguish between the two. This example however, illustrates only one out of the many applications of simulation in a large-scale HEP experiment such as the LHCb. In addition to the calculation of the detector efficiency, simulation is also involved in the very early stages of the experiment to design and optimize the detectors for best physics performance, as well as in the calculation of the performance of the detector which is crucial for the extraction and interpretation of the available data (indeed, the numbers provided at the beginning of Section 4.3 regarding the performance of the LHCb detector can only be estimated by simulation). It is also heavily used for the estimation of background signal in the extracted data and the evaluation of the possible physical processes in the various parts of the sub-detector in order to assess their impact on the final data model via the calculation of uncertainties.[25]

For the purposes of our discussion, it therefore suffices to say that although what counts as simulation data and what counts as signal data in the experiment is quite straightforward, the data that reaches the experimenters as reconstructed events for further analysis, is in effect an indistinguishable amalgamation of these two types. Along with a number of additional factors, simulation data therefore have a clear influence on the final properties of the data model either directly via their presence in the processed datasets that reach the scientists' desks, or indirectly via their effects on the various aforementioned stages and procedures of the experiment. Although the discovery of a new particle or the presence of new physics in a physical process cannot – of course – be claimed based solely on simulation data, the final data model that is eventually compared to theory to make such claims is in effect a *co-production* of real and simulated data. The extent to which each type of data contributes to the final results depends on the specific details of the experiment. This further suggests that the question whether real data are more reliable than simulated data does not really apply in the case of HEP, since in practice, there is rarely a case in which a dataset is exclusively constructed from real data.

4.6 Evaluating the results

The description of the four stages in HEP data modelling and the following remarks on the two distinctions between raw/processed data and real/simulated data bring us to the end of this chapter. As we have seen, the construction of a data model in HEP

---

[25] For a detailed review of the impact of simulation to collider experiments in general, including the discussion of cases where the use of simulation samples made a difference in the precision of the physics results, see Elvira (2017). For a full description of the LHCb simulation system see Clemencic et al. (2011).

typically proceeds via a four stage process in which (i) the selection criteria for reducing the available data are defined and applied at the trigger systems, (ii) the efficiency of the detector in recording the relevant events is calculated, (iii) the yields of the decays are determined by data fitting, and (iv) the uncertainties accompanying the final result are determined and calculated. It is important to note once again, that in practice, these four stages are not clearly separated during the data analysis, nor do they follow a linear path in which one stage follows after the completion of another. Rather, the activity of constructing a data model is a long and iterative process of trial and error, in which several attempted algorithms for extracting the result go back and forth a peer review process until they reach the necessary standards for publication. The breakdown of these procedures in four different stages only aims in giving an overview of the main tasks that need to be accomplished in order to compress the huge amount of information hidden in the available data into a simple data model to be compared with the theory.

Fig.4.4 perfectly captures Suppes' dictum that theoretical hypotheses are eventually compared with 'a simple entity' – i.e. the data model – which incorporates all the relevant information extracted from the many and various procedures that constitute the LHCb experiment. The graph is taken from the LHCb's most recent announcement of the result, and shows the comparison between the theoretical prediction of the Standard Model (vertical dashed line) and various experimental results for the R_k ratio (horizontal lines). Compared to the previous result at the LHCb (LHCb collaboration, 2019) and the results from the Belle and BaBar experiments, the 2021 LHCb result by far has the smallest associated uncertainty which makes it the most precise and robust measurement of the R_K ratio to date.

This result is consistent with the Standard Model prediction at the level of 3.1 standard deviations, which corresponds to a p-value of $\sim 0.1\%$. In practice, this means that if the 'null hypothesis' is correct – i.e. if there is no violation of Lepton Flavour Universality – then the probability of obtaining any data yielding a discrepancy from the Standard Model prediction that is at least as great as that obtained with these data is about 1 in 100000. The 3.1 sigma level is still far away from the golden 5 sigma level for claiming a new discovery in particle physics, corresponding to the much lower p-value of approximately 1 in 3.5 million. This, however, is a significant improvement to the 2.5 standard deviation of the 2019 measurement with a p-value of 1 in 166, in that it comes with even smaller uncertainties and makes the possibility of discovering new physics in rare B-decays more credible. Future measurements of the ratio based on larger data samples are expected to both reduce the total uncertainty and increase the

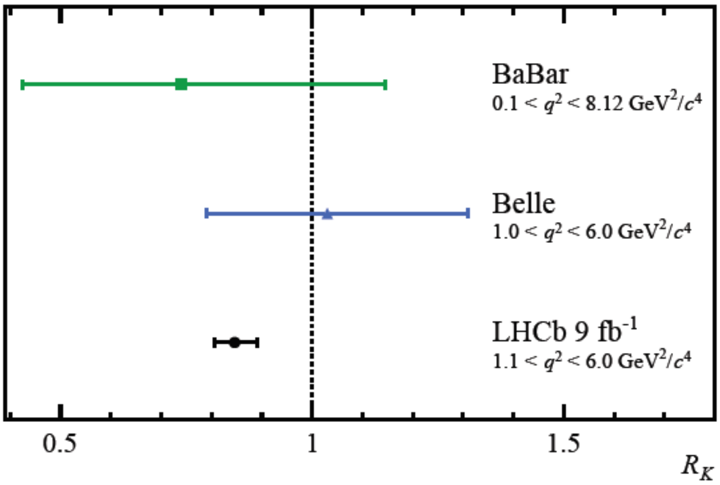

Figure 4.4: Comparison of the Standard Model theoretical prediction and various R_K ratio results from different experiments. (LHCb collaboration, 2021).

sigma level in order to reach a more definite conclusion for the possible violation of LFU in B-decays.

4.7 Conclusions

To summarise, the main objective of this chapter was to explore the connection of theory with experimental results with the use of data models, by studying in detail an example of experimental practice in HEP. Our discussion began with a brief presentation of Suppes' hierarchy of models account and his distinction between models of theory, models of experiment and models of data. The following section focused in providing the theoretical framework of the rare decays of B-mesons at the LHC in order to understand the experimental process of LFU tests at the LHCb for which the data model of the R_K ratio is constructed. The discussion continued with a presentation of the LHCb trigger system, followed by the presentation of the four main stages for the construction of the data model of the R_K ratio and some remarks on the two distinctions between raw/processed data and real/simulated data in support of the four main conclusions of this chapter. Finally, we have seen how the final data model is eventually evaluated and compared with the corresponding theoretical model.

The first conclusion is that the first data collected at the early stages of the experiment, which can be characterised as the raw data of the experiment, are useless as they are for the comparison between theory and experimental results, since they necessarily need to undergo a process of refinement in order to be transformed into a language that is comparable to theory. This also indicates that raw data in HEP cannot be understood in the traditional sense as data directly perceived from human experience and that, contra to popular perception, the process of refining the data sometimes makes the processed datasets epistemically more reliable than non-processed data.

The second conclusion concerns the fact that the final datasets that reach the hands of physicists for analysis consist of a mixture of simulated and real data that cannot be distinguished. The use of simulation and its data are essentially involved directly or indirectly in almost every step of the data acquisition and data analysis process and hence, one can safely say that the final data model of the R_K ratio that is eventually compared to the theoretical prediction of the Standard Model is a co-production of data coming from the physical interactions of particles in the detector and computer simulations.

The third conclusion is that theory guides the observation and the derivation of results in three different levels: a fundamental level which is universal across all experiments in HEP, an intermediate level regarding the various processes throughout the experiment which are not directly involved with the physical phenomenon under investigation, and a third and most specific level which explicitly guides the overall experimental procedure based on the specific research question of the experiment. In the core of these three levels lies the fundamental assumption that new physics will resemble known physics. This means that the anticipated models and theories that go beyond the Standard Model are expected to respect all the fundamental laws of current physics, and new physics will only appear in extremely short distance/high energy scales and in rare processes such as the decays of B-mesons which have not yet been studied in detail.

The fourth conclusion is that Suppes' categorization is not as rigid as one might first think, in that the three types of models cannot always be easily distinguished. Nonetheless, this categorization remains a useful conceptual tool for describing the otherwise extremely complicated structure of large scale experiments in HEP. In this context, a data model can be understood as the representation of an experimental result in the form of a graph, table or numerical answer that allows the comparison of experiment with theory. While this straightforward answer to the question of what a data model is does not differ from what Suppes and others have said, what is of special philosophical interest is the complicated and extremely laborious process of constructing a data model in HEP, which has largely been overlooked by philosophers of science. The detailed analysis of the necessary considerations regarding the determination of cuts, the calculation of efficiencies and uncertainties and the fitting of data with sophisticated algorithms shows that the process of constructing a data model in HEP involves much more than the mere collection and organization of raw data, and cannot be as easily formalized as Suppes implied.

In addition to these main conclusions, the detailed description of the idiosyncrasies of the LHCb experiment for the test of LFU and the various challenges faced by physicists in their attempt to derive the experimental results also reveals a number of further issues worthy of philosophical attention. The pragmatic dimension of the experimental process regarding the determination of selection criteria based on time limitations, computational power and store capacity, and the fact that the LHCb detector is able to collect only 25% of the b-quarks that are produced in the proton collisions of the LHC means that the otherwise huge amount of data that eventually gets stored for further analysis is only a tiny fraction of the potentially available data from the proton-proton collisions in the LHC. Although special attention is given to collect the most relevant data with respect to a research question, it is a widely accepted fact that the data that are irretrievably thrown away at the LHC contain evidence for new physics and hence, the final data model of the R_K ratio, as well as most of the results in HEP, is not a solid and flawless representation of reality in the microscopic scale as one might think. Rather, it is itself a hypothesis based on our best estimation given the small fragment of data we are able to collect from particle collisions.

As a final remark, let us also note that the plethora of experimental results showing potential anomalies at the 2-3 sigma level has already led to the development of various phenomenological models containing new physics in the form of additional interactions that allow the violation of LFU. The most promising types of such models involve the existence of additional particles such as the so-called 'leptoquarks' (Becirevic et al. 2016) or a new heavy neutral Z' boson (Celis et al. 2015). This fact nicely reflects our observations in Chapter 1 regarding the driving force for the construction of phenomenological models in physics. Despite the fact that the violation of LFU has not been decisively established, the existence of hints at the 3-sigma level, combined with other existing results indicating potential disagreements with the Standard Model, suffices to initiate the construction of new phenomenological models with the aim of eventually refining or even replacing the existing theory, indicating once again that the art of modelling often lies at the core of the development of new theories in physics. The precise way in which indirect searches in HEP, such as the measurement of the R_K ratio, give rise to new phenomenological models extending the Standard Model of particle physics, and the impact of these models on future research in HEP is an interesting topic that deserves to be explored further in future work. In the next and final chapter, we shall turn our attention to the domain of astrophysics, and examine the various challenges in the ways in which the phenomenological models of dark matter can be tested based on data from the various methods of dark matter observation.

Chapter 5

The Puzzle of Dark Matter Observation

According to the current received view in cosmology, the ΛCDM model, more than 95% of the observable universe is 'dark', consisting primarily of dark energy (≈68%) and dark matter (≈27%). However, despite the fact that the systematic research of dark matter has a history of almost fifty years, its exact nature still remains elusive mainly due to the severe underdetermination of viable dark matter models by the available evidence. If it exists, dark matter could be anything insofar as it satisfies a minimum set of constraints based on current cosmological observations. Martens (2022) describes this minimum set of constraints as the 'thin common core concept of dark matter': assuming standard physics, dark matter is a massive field that contributes $\approx 27\%$ to the total cosmic mass-energy budget, it primarily interacts with baryonic matter via the gravitational force, and if it is a particle, its mass is expected to be between $10^{-3} - 10^{7}$ eV (Calmet & Kuipers, 2021). If one is willing to give up the standard gravitational laws of physics, the thin common core becomes even thinner: dark matter is some sort of 'stuff' that contributes $\approx 27\%$ to the total cosmic mass-energy budget or acts as if it does so, and is responsible for the observation of various 'dark phenomena' related to structure formation, clusters and galaxies.[1]

---

[1] The term 'dark phenomena' is borrowed from Martens & Lehmkuhl (2020a) and refers to the various astrophysical phenomena that either contradict the gravitational laws of General Relativity or require the postulation of additional 'invisible' dark matter that causes the formation of some large scale cosmological structures due to its gravitational pull. Examples of such phenomena are the mass discrepancies in the Coma cluster and the flat rotation curves of nearby galaxies. For a nice review of the observational evidence for dark matter and dark energy based on dark phenomena see Jacquart (2021).

The elusive nature of dark matter and the various methodological conundrums faced by particle physicists and cosmologists in dark matter research has attracted moderate attention by philosophers (e.g. Vanderburgh 2003, 2005, 2014; Kosso 2013; Massimi 2018; Weisberg et al. 2018; de Swart 2020; Martens & Lehmkuhl 2020a, 2020b; Smeenk 2020; de Baerdemaeker 2021). However, the existing body of philosophical works remains largely unintegrated and cannot yet be considered as a cohesive and interconnected philosophical literature on the problem of dark matter. The aim of this chapter is two-fold. First, it provides a common ground for the discussions on the philosophy of dark matter by offering a novel framework for the epistemology of direct and indirect dark matter observation and a presentation of the five possible methods for dark matter observation based on the available phenomenological models for the nature of dark matter (Sections 5.1 and 5.2). Second, it is argued that a partial explanation for the lack of progress in dark matter research despite the huge amount of invested time and funds during the last decades, stems from the fact that robustness arguments from the variability of experiments are significantly limited within the results of methods that are based on the same models (Section 5.4). Progress in dark matter research is to be understood here as the elimination of the viable phenomenological models of dark matter by enriching the common set of constraints from which these models are built, in a way that allows physicists to learn more about the nature of dark matter. The main argument of the chapter is supported by an evaluation of the 'epistemic strength' of the various methods for dark matter observation, based on their *informativeness*, *model sensitivity*, and *reliability* (Section 5.3).

Robustness arguments are widely used in various forms both from scientists and philosophers.[2] For the purposes of this chapter, *robustness from the variability of experiments* will be understood in terms of what Woodward (2006, p.233-5) calls 'measurement robustness'. The rationale behind measurement robustness is that if different measurement procedures of the same quantity Q that are in some sense independent of each other produce nearly the same result, then the result is said to be *robust* and can be used as grounds for increasing our confidence that the quantity has been measured accurately. This is because it is very unlikely that each procedure is subject to exactly the same kinds of error that would give rise to the same result, and so we have good reasons to believe that the result of the different measurements is reliable.

This characterisation of measurement robustness by Woodward is closely related to

---

[2] See for example Franklin (1989), Weisberg (2006), Woodward (2006), Kuorikoski, Lehtinen & Marchionni (2010), Parker (2011), Odenbaugh & Alexandrova (2011), Lloyd (2015), Eronen (2015), Basso (2017), Lisciandra (2017), and Schupbach (2018).

what Franklin (1989, Ch.6) describes as the *variability of experiments*, as a possible strategy to justify the reliability of experimental results in high-energy physics. For Franklin, a possible agreement between the experimental results of two or more different experimental methods automatically increases our confidence in the reliability of the results, since it would be a 'preposterous coincidence' if two different physical systems produced the same false results. On the surface, given the high competition between viable dark matter models and the existence of different methods for probing the properties of dark matter and constraining the parameter space of these models, such arguments of robustness from the variability of experiments can be applied in dark matter research in favour of those model hypotheses and properties that are supported by the results of more than one experimental method. If multiple methods provide the same results about the constraints of one or more model parameters, then this parameter should be considered as more robust, compared to the parameters that can be measured by only one method.

In a similar spirit, Stegenga (2009) defines robustness as 'the state in which a hypothesis is supported by evidence from multiple techniques with independent background assumptions' (*ibid.*, p.651). The main idea behind Stegenga's definition is that hypotheses are better corroborated by results coming from multiple and independent experimental methods compared to hypotheses that are supported by the results of a single experimental method. It is not hard to see how this is related to Woodward's measurement robustness and Franklin's variability of experiments: given that a hypothesis can be understood as encompassing the scientists' belief about the value of a specific model parameter or the allowed parameter space for one or more parameters of a model, the scientist's confidence about the truth of these hypotheses is increased if they are supported by more than one experimental method.

Stegenga's aim is to argue that while robustness is a valuable epistemic guide in 'ideal epistemic circumstances', when it comes to real scientific practice it faces important limitations. For Stegenga, the main problem with robustness arguments is that, in practice, most of the time multiple and independent experimental techniques provide results that are inconsistent (i.e. one method suggests x and another method suggests $\neg x$) or incongruent (i.e. one method suggests x and another method suggests y), and hence, it is not clear what kind of epistemic support is provided to the relevant hypotheses.[3] While this seems to be the case in dark matter research as well, it will be argued that even in cases where multiple methods ostensibly provide concordant results on the constraints of certain parameters, the dependence of such results on a number of

---

[3] See Lehtinen (2013) and Hey (2015) for a response.

factors concerning the introduced assumptions in the experiments and the dependence of the experiments on specific dark matter models, makes the task of establishing that the results of different experiments are actually the same extremely difficult. Hence, robustness arguments from the variability of experiments face important limitations in dark matter research, not only because the results of multiple methods rarely agree, as Stegenga notes, but also because even in cases of a potential agreement on the surface, establishing the concordance of these results is challenging.

The upshot is that, at best, arguments from the variability of experiments can only be applied – with caution – for the comparison of hypotheses concerning the *sub-models* of a broader model scenario for dark matter. That is, the corroboration of a model specific hypothesis within a model scenario of dark matter via multimodal evidence is a good epistemic guide for preferring a specific sub-model over another, but it says nothing about the prevalence of the relevant model scenario over alternatives. For instance, arguments from the variability of experiments have epistemic importance only when applied for the comparison of different WIMP sub-models (e.g. the Lightest Supersymmetric Particle (LSP) and the Lightest Kalusa-Klein particle (LKP)), but are practically irrelevant for the comparison between, say, hypotheses concerning the competing scenarios of dark matter as WIMPs or as axions.[4]

This important limitation highlights the difficulty of narrowing down the range of viable phenomenological models for the nature of dark matter in order to make progress in dark matter research. To learn more about the nature of dark matter is to reduce the large number of possible models to a single ‘true’ model, or a set of true models that are supported by observational evidence and adequately explain the various dark phenomena in astrophysical observations. This can be achieved by enriching the common core concept of dark matter via the introduction of further constraints which will eventually preclude certain model scenarios for the nature of dark matter and will favour others. However, it seems that the only possible way to gather more reliable information via the various observational methods of dark matter in order to enrich the common core is by presupposing the very same models for which there is no observational evidence in the first place and for which no robustness arguments can be used. This complication is what I shall call the puzzle of dark matter observation.

The present chapter aims in making fruitful contributions to the discussions of robustness arguments from the variability of experiments and the broader philosophical

---

[4] The LSP (also known as the neutralino) and the LKP are probably the two most well-studied sub-models for WIMPs, coming from supersymmetry and theories of extra dimensions respectively. The terms ‘model scenario’ and ‘sub-models’ will be clarified further in the next section.

literature on dark matter. By providing a conceptual analysis of dark matter observation and taking a close look at current scientific practice, it brings to light an important problem faced by scientists concerning the limits of robustness arguments for the confirmation of results coming from different methods. It also illustrates a broader concern in experimental physics, which can probably be found in other areas of science as well, stemming from the fact that meaningful results can be extrapolated from experiments only on the basis of certain assumptions that are not necessarily supported by empirical evidence. As a result, a growing number of model specific constrains constantly find their way in the literature, however, the essential question about which model or models captures the actual nature of dark matter remains unanswered. While it is often asserted that the slow progress in dark matter research is due to the lack of positive results in direct, indirect and collider searches, the present study illustrates that even in the fortunate event of a dark matter particle discovery by one or more of these methods, there is still a long way until the exact nature and the properties of such particle are determined.

At the same time, the following analysis in this chapter highlights the importance of phenomenological models in providing how-possibly explanations in physics, as discussed in Chapter 1. The fact that certain dark phenomena cannot be accommodated by the present theoretical framework of the standard model of particle physics and general relativity, initiates the construction of several phenomenological models of dark matter aiming to accommodate the available evidence for dark matter providing how-possibly explanations. Moreover, as we shall see, the abundance of viable models of dark matter nicely illustrates how models can contribute to the aim of the economy of thought, by classifying and condensing several models into groups of simplified models capturing their essential and common components. For instance, by classifying various models with specific characteristics as models of Weakly Interactive Massive Particles (WIMPs), scientists are able to identify the common assumptions throughout these models and constrain their parameter space by conducting a single type of analysis on the available data.

5.1 Observing dark matter

Before we delve deeper into the analysis and evaluation of the various methods for dark matter observation, it is useful to clarify how the term ‘observation’ should be understood in modern physics, and draw two useful distinctions in order to better understand the various ways in which dark matter can be observed.

Within the scientific context, observation can be defined as the act of obtaining information about the properties of one or more physical entities, via any kind of interaction which involves the communication of information from the target of the observation to the observer. This definition closely follows Shapere's (1982) understanding of observation as a subspecies of interaction between two physical systems where information is transmitted from one system to another. Peter Kosso (1989, p.30) also gives a similar definition of observation as 'a manner of getting information of the physical world, from the physical world' which is 'accomplished with the conveyance of information from the world to the scientist, information like "x is P", where x and P are features of the physical world'. In a sense, observation is thus a binary directional relation between the target system(s) being observed – i.e. the physical entities about which the scientific community is gathering information – and an observer. The observer can be either the scientist or any other scientific device which can be used for the collection of information about the properties of a physical entity, which will eventually be conveyed to the scientist via a series of data.

This definition is of course different from the traditional understanding of observation as the act of seeing something or, more general, the act of perceiving information about something via the human sensory system. Nonetheless, it has been carefully chosen to reflect the application of the term in modern scientific practice (especially in high-energy physics and cosmology), and will be used as a working definition for the purposes of our discussion. Following Shapere's and Kosso's approach to the study of observation in science, the aim here is not to define what observation *really* is, but rather, to examine the act of observation as an epistemic event in physics in order to shed light on what actually happens when scientists talk about the various ways in which dark matter can be observed.

The first useful distinction for our discussion is the one between *mediated* and *unmediated* observation. As already mentioned, when information is conveyed from a target system to an observer, the observer acts as a receptor of this information. If the immediate receptor of this information is the scientist, that is, if the existence of a physical entity and the relevant information about its properties are directly perceived by the human sensory system without the use of an auxiliary device, then one might talk of unmediated observation. Alternatively, if the relevant information about a target system is collected with the help of a mediating system whose aim is to collect, store or transmit the information received via its interaction with the target, then the observation becomes mediated. Quite often, this mediating system is what Humphreys (2004) calls an *epistemic enhancer*, i.e. a scientific device that facilitates our access to information which would otherwise be unavailable to us. This is achieved by the ability of these objects to interact with certain physical entities in ways that are, *in principle* or *in practice*, unavailable to humans.

Notice that the distinction between mediated and unmediated observation is a distinction concerning the *act* of observation and is not directly analogous to the observability of physical entities. That is, while an, in principle or in practice, unobservable entity can only be observed via a mediated observation, the opposite is not true. It is always possible, for instance, to use a telescope for the observation of the surface of the moon, but this does not entail that the moon is an unobservable physical entity. Evans and Thébault (2020, Sec. 2(b)) offer a very illustrative three-fold categorization of unobservable entities as (a) manipulable, (b) unmanipulable accessible and (c) unmanipulable inaccessible. The first category concerns those entities for which there is 'two way' causal access, in that we can probe the phenomena via a suitable mediating system, and the phenomena can somehow 'respond' by providing information. Unmanipulable physical entities on the other hand, are those entities whose behaviour is beyond our experimental control and thus the nature and the amount of information we can receive from them is limited. If we have access to any sort of information about an unmanipulable entity (e.g. measuring gravitational waves from a black hole merger), then it is accessible. If not (e.g. a physical entity that lies beyond our cosmological horizon), then it is inaccessible to us. As we shall see, depending on which model of dark matter eventually corresponds to the physical world, it is possible that dark matter belongs to any of the aforementioned three categories. Nevertheless, what is important to keep in mind, is that dark matter – if it exists – belongs to the realm of the unobservable world – as most of the subjects of modern particle physics and cosmology do – and thus, its observation necessarily has to be mediated.

The second useful distinction for our discussion is the one between *direct* and *indirect* observation. An observation of a physical system is direct if the target of the observation directly interacts with – what/who is considered to be – the observer. Recall that the act of observation, as defined earlier, concerns the conveyance of information from the target of the observation to the observer, regardless of whether the observer is a human or an artificial receptor of information such as a scientific measuring device. If the reception of information is a result of the direct interaction between the target and the observer, then the observation is direct. Conversely, an observation is indirect when the physical entity interacting with the observer, is *not* the target itself, rather, it is a physical entity which

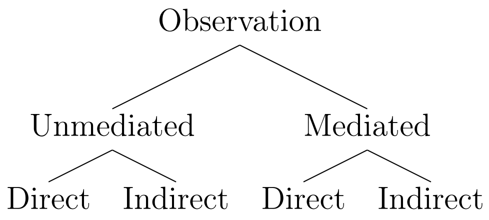

Figure 5.1: Four modes of observation

is believed to be causally related to the target. The main idea behind indirect observation is that there is a causal series of events whereby the target interacts with other physical entities which in turn interact with the human observer or with an artificial receptor producing recognizable effects that are subject to scientific interpretation.

When combined, the two distinctions between mediated/unmediated observation and direct/indirect observation give rise to four different modes of observation (Fig.5.1). By means of examples, an unmediated direct observation would be the seeing of an aeroplane in the sky with the naked eye, whereas an unmediated indirect observation would be the observation of an aeroplane via its contrails. On the other hand, an example of mediated direct observation would be the famous observations of the phases of Venus by Galileo. What makes these observations mediated is the fact that Galileo was only able to observe the various phases with the use of his telescope, and what makes them direct is the fact that the physical entity interacting with the observer was the target itself (i.e. Venus) via the transmission of light. An example of a mediated indirect observation comes from the celebrated observation of the Higgs boson at the LHC in 2012. Due to its very short lifetime, the Higgs boson decays too quickly to interact with the detectors of the LHC, and hence, its existence can only be inferred by analysing the various products of the proton collisions at the LHC, in order to identify a decay signature of a Higgs boson from which the production of the Higgs boson can be inferred. The observation of the Higgs boson is therefore mediated in that it can only be accomplished with the help of a particle detector, and it is also indirect in that the entity which directly interacts with the various sub-detectors of ATLAS and CMS at the LHC, is not the Higgs boson itself, but the decay products that are causally related to it.

It should be clear by now that whether a mediated observation is classified as direct or indirect depends on the arbitrary choice of where to draw the line between the observer and the target. That is, the very same physical process in which a series of causally related events eventually produces a dataset, can both be described as a direct or indirect observation depending on what is eventually taken to be the target and the observer amongst the involved physical entities. Given the arbitrariness of this choice, one may wonder why this distinction is useful at all, however, this is a very important point which will become clearer later in the description of the five possible methods for dark matter observation. The two distinctions made here have been drawn in order to reflect the distinction between direct and indirect searches of dark matter made by the physics community. As will be shown in the next section, the clear articulation of these two distinctions serves in solving an apparent confusion regarding the fact that physicists often refer to the direct observation of dark matter, although this method of observation seems to be, in a sense, highly indirect. It also shows how the indirectness of an observational method is ultimately a matter of degree depending on how long the involved series of causal events is.

5.2 Many models, five methods

The fact that the nature of dark matter is severely underdetermined by the available cosmological evidence has naturally led to the development of a large number of phenomenological models for dark matter. Some of these models concern the particle nature of dark matter (e.g. purely collisionless dark matter, self-interacting dark matter, dark matter as weakly interacting massive particles (WIMPs), sterile neutrinos, axions etc.), while others describe the large scale structure of dark matter in terms of Massive Compact Halo Objects (MACHO's) which may or may not compose of baryonic matter, such as primordial black holes (Carr et al., 2016), and neutron stars (de Lavallaz & Fairbairn, 2010). As already mentioned, dark phenomena are astrophysical anomalies which either contradict the fundamental gravitational laws of general relativity or can only be explained by postulating the existence of undetected dark matter. Given that the existence of dark matter particles is not a consequence of the standard model of particle physics, the aforementioned models follow the second option by extending or modifying the standard model in order to provide a how-possibly explanation for these anomalies. In other words, they are phenomenological models that often contradict central principles of the background theory – i.e. the standard model – in order to accommodate the available experimental data from the observed anomalies, that essentially define the common core concept of dark matter. Nevertheless, a large part of dark matter research is also devoted to the development of alternative gravitational models and theories in which no postulation of dark matter particles is required. The most prominent example comes from the models of Modified Newtonian Dynamics (MOND) in which Newton's inverse square law of gravity is modified accordingly in order to reflect astrophysical observations on the galactic scale and also account for the observed rotation curves of individual galaxies (Sanders & McGaugh, 2002). Another possible way to explain dark phenomena is to see dark matter and dark energy as low energy quantum gravity effects of a quantum gravity theory that approximates general relativity in the limit, but differs from it in the fine details (Reichert and Smirnov, 2020).

For the remaining of this chapter we shall refer to all frameworks for the possible explanation of dark phenomena as the viable *model scenarios* of dark matter, whereas the various specific examples of a particular model scenario – e.g. the LSP and LKP examples in WIMPs – will be referred to as the *sub-models* of a model scenario.[5] The reason behind the choice of this particular terminology is because some of these model scenarios – e.g. WIMPS – are blanket terms for a group of sub-models that share a common set of characteristics, while others – e.g. MOND – are , in essence, mini-theories whose various models can accommodate the data from dark phenomena.

Going back to our discussion in Chapter 1, the phenomenological models of dark matter nicely illustrate how their construction aims in achieving the three outcomes we have discussed, namely the economy of thought, the explanation of data, and the prediction of physical phenomena. First and foremost, as already mentioned in the first chapter, model scenarios of dark matter are primarily constructed with the aim of providing how-possibly explanations for the nature and behaviour of large scale structures in the local Universe which cannot be accommodated by the standard model of particle physics and general relativity. The aim is to construct phenomenological models which will eventually turn into theoretical models via their embedding in a well established theoretical framework. The new theoretical framework can be an extension of an existing theoretical framework – e.g. in the sense that supersymmetry featuring the neutralino particle is an extension of the standard model – or a replacement of a current well-established theory – e.g. in the sense that a fully fledged version of Relativistic MOND aims in replacing general relativity.

Moreover, the grouping of various candidate models in terms of their common fea-

---

[5] Just to provide a sense of how long the list of dark matter models is, some further proposed dark matter candidates that are not mentioned above are: gravitinos, axinos, light-scalar dark matter, dark matter from little Higgs models, wimpzillas, Q-balls, mirror particles, CHArged Massive Particles(CHAMPs), Strongly Interacting Massive Particles (SIMPs), D-matter, cryptons, superweakly interacting dark matter, brane world dark matter, heavy fourth generation neutrinos etc. The possible masses for these candidate particles for dark matter extend over 90 orders of magnitude (Bertone & Tait, 2018). For a generic review of dark matter research containing useful information on various dark matter candidates see Bertone et al. (2005). For a detailed review of dark matter candidates see Feng (2010).

tures nicely illustrates a possible way in which the economy of thought can be achieved in physics. For instance, as already mentioned, the term Weakly Interactive Massive Particle (WIMP) is a blanket term for models of dark matter featuring particles that interact with ordinary baryonic matter via gravity and a non-vanishing force which is either weaker or at least as weak as the weak nuclear force. By classifying a large class of models as WIMP's, scientists are able to construct various *simplified models* that aim in producing a common prediction which depends on a shared set of assumptions amongst the different but similar WIMP models for dark matter. This practice essentially allows the simultaneous test of a multitude of models for WIMPs, since the refutation of the common prediction from the simplified models would imply the unsuitability of the entire class of models from which the simplified model is built, without the need for a detailed and rigorous development of the mathematical framework of each model.[6]

As for prediction, a large part of the debate between advocates of alternative theories of gravity and advocates of dark matter existence lies on the predictive success of the corresponding models in the relevant model scenarios. In particular, proponents of MOND often appeal to the success of MOND models in predicting several theoretical relationships between various observable quantities in the galactic scale, such as the Tully-Fisher relation between the mass of a spiral galaxy and its asymptotic rotation velocity (e.g. Lelli et al. 2016; McGaugh 2020). On the other hand, dark matter models that maintain general relativity predict the existence of galaxies lacking dark matter, an extremely rare phenomenon which was recently discovered by cosmological observations (van Dokkum et al. 2018). One might therefore say that the various phenomenological models for the explanation of dark matter phenomena aim towards making empirical predictions that will confirm some models and disfavour others, eventually leading to a complete theoretical framework which will accommodate the currently observed anomalies related to dark phenomena in their entirety.

As one might expect, the abundance and diversity of viable model scenarios for dark matter has naturally led to the development of a variety of methods for the possible observation of dark matter, since each model scenario is built on a number of different assumptions and requires different experimental setups to be tested. In particular, dark matter can be observed according to current scientific practice via five different methods based on the physical phenomena on which they rely: (a) via its gravitational effects (b) via cosmological observations (c) directly (d) indirectly and (e) in collider searches. Fol-

---

[6] For more on the nature and the importance of simplified models in high-energy physics see McCoy and Massimi (2018).

lowing our discussion in Section 5.1, the observation of dark matter is understood here in the scientific sense of acquiring information about the existence, the behaviour and the properties of dark matter in various scales. Given that dark matter is non-luminous and presumably nonreactive, it should also be clear that it can only be observed in a mediated way, i.e. via its (direct and indirect) effects on our measuring instruments. Finally, while the observation of dark matter based on gravitational effects and cosmological observations is often considered as providing strong evidence for its existence, it should be noted that none of the remaining three methods have offered any significant results so far, other than constraining the parameter space of various models. Below is a brief description of the five possible methods of dark matter observation.

Gravitational effects: If dark matter consists of particles which primarily interact with baryonic matter via the gravitational force, it is possible to observe its gravitational effects on large scale structures of the observable Universe. Some of these effects are the mass discrepancies in galaxy clusters observed by Jan Oort (1932) and Fritz Zwicky (1933) in the early 1930's, as well as the flat rotation curves first studied by Kenneth Freeman (1970) and Vera Rubin and Kent Ford (1970). An additional way of observing dark matter by its gravitational effects is via the phenomenon of gravitational lensing. Roughly speaking, gravitational lensing is an implication of general relativity, in which the images of luminous objects in the background are distorted by the presence of massive objects in the foreground as a result of the gravitational pull of light. In the literature on dark matter, one may find a plethora of lensing systems where invisible dark matter clumps are considered to be responsible for the distorted images of various celestial objects in our telescopes.[7]

Cosmological Observations: Another method that is considered to provide strong evidence for the existence of dark matter comes from precision measurements of various observables on the cosmological scale. The central goal of these methods is to track down any possible non-gravitational effects of dark matter interactions on the large-scale structures and the thermal history of the universe. These results can then be used to probe the underlying physics of dark matter by placing various bounds on its properties. The main underlying assumption is that the elastic scattering between dark matter and baryonic matter in the early universe leads to an exchange of heat and momentum between dark matter and baryons, which in turn affects the thermal history of the Universe and the evolution of cosmological perturbations. These effects are captured by various

---

[7] A widely discussed example of a dark matter gravitational lens comes from the Bullet cluster of galaxies which consists of two galaxy systems that have traversed one another (Clowe et al. 2006). For a review of dark matter gravitational lensing see Massey et al. (2010).

cosmological observables related to the Cosmic Microwave Background (CMB) such as its spectral distortions, polarization, and temperature anisotropies.[8] Other cosmological observables providing information about the nature of dark matter are related to data from distance measurements of Type Ia Supernovae (SN Ia) and Baryon Acoustic Oscillations (BAO). These measurements are instrumental for the determination of the late time expansion rate of the universe that can be explained by the contribution of dark energy, however, the data also constrain the value of the total dark matter density.[9]

At the time of writing, the most accurate data on cosmological observables related to the above searches come from the Planck mission (Tauber 2004), a space observatory operated by the European Space Agency between 2009 and 2013 that has improved the previous observations made by the NASA's Wilkinson Microwave Anisotropy Probe (WMAP). The combination of datasets from CMB observations with BAO and SN Ia measurements provides a precise estimate of the average matter density in the Universe, as well as a tight constraint on the mass of the dark matter particle (Aghanim et al. 2018).[10]

Direct Searches: Direct detection experiments are based on the interaction of dark matter particles with ordinary baryonic matter. The basic idea behind direct searches is that if the galaxy is full of dark matter particles (e.g WIMPS, axions etc.) that interact weakly with baryonic matter, then a significant amount of them will travel through the Earth, enabling us to search for the interaction of these particles with Standard Model particles by recording the recoil energy of nuclei as dark matter particles scatter off them. When (and if) these interactions take place, two types of processes are expected to happen. The first and most common process is the elastic (WIMP-nucleus) scattering in which the dark matter particle interacts with the nucleon as a whole. The second and less common process is the inelastic (WIMP-electron) scattering in which the dark matter particle interacts with orbiting electrons of the target material either exciting them or ionizing the target. In this case, the scattering is not observed by recording the

---

[8] To provide a short example, the frequency spectrum of CMB has an almost perfect black-body form containing some very small – but measurable – anisotropies. These anisotropies can be decomposed into an angular power spectrum consisting of a series of acoustic peaks and troughs. The shape of the spectrum implies – amongst other things – that about 27% of the total mass density of the universe is dark matter.

[9] Other types of precision measurements related to 21-cm cosmology and Lyman-a forest observations are also able to provide constrains on the nature of dark matter, however, these measurements are still difficult to make and currently suffer from large uncertainties.

[10] For more detailed reviews on the cosmological observations of dark matter see Lukovic et al. (2014) and Gluscevic et al. (2019).

nuclear recoil, rather the process also involves the emission of a photon.[11]

Since the interaction of dark matter particles with the nuclei is expected, by definition, to be extremely weak, direct search experiments take place in ultra-sensitive low-background experiments that are often placed well below the Earth's surface in order to block out spurious particles. An example of an experiment set up to search for the possible rare collisions of dark matter particles coming from the Milky Way halo is the Large Underground Xenon (LUX) experiment (Akerib et al. 2013) at the Sanford Underground Laboratory in Lead, South Dakota. The experiment comprises a large tank of liquid xenon located 1.5 km underground, surrounded by ultra-sensitive photomultiplier detectors.[12] When a particle interacts with the nuclei of xenon it produces heat along with ionised atoms. The excited atoms in turn combine with the neutral atoms of the gas forming excimer states[13] which subsequently decay under the emission of detectable ultraviolet light.

This process is a prime example of the *causal chain of events* that take place in a mediated observation. The 'directness' of these experiments has nothing to do with the fact that the target of the observation is directly perceivable by the observer. Rather, as already shown earlier, it indicates that the target of the observation – i.e. the dark matter particle – directly interacts with our detector, which in the case of the LUX experiment is the liquid xenon, producing a detectable signal which leads to a series of data.

Indirect Searches: Indirect searches for dark matter are based on the astronomical observation of Standard Model particles that are most likely to be the products of the decay or annihilation of dark matter in the Universe. These searches are based on the assumption that the final states of a dark matter annihilation/decay are either Standard Model particles of any kind (insofar as they are kinematically accessible), or unknown particles which then decay to Standard model particles. Current experiments for the indirect detection of dark matter are mainly focused on the detection of three different products: (i) gamma-rays, (ii) neutrinos and (iii) cosmic rays.[14]

These three types of radiation are used for the indirect detection of dark matter for a number of theoretical and practical reasons, such as the fact that the mass scale of WIMPs in the most promising models implies that a large fraction of the generated

---

[11] See Schumann (2019) for a review of direct searches on dark matter.

[12] Noble gases such as argon and xenon are commonly used as WIMP detectors since they are excellent scintillators and can be ionized easily.

[13] An excimer (short for Excited Dimer) is a short-lived molecule consisting of two species, at least one of which is electronically excited.

[14] See Gaskins (2016) for a detailed review of indirect searches of dark matter.

emission from dark matter annihilation/decay ends up in gamma-ray energies. The techniques used to detect the signals from these products vary. For instance, gamma-rays travelling from distant regions of the universe are mainly absorbed by the Earth's atmosphere due to their small wavelengths, and therefore, their detection can only be achieved by space based telescopes such as the Fermi Large Area Telescope (Acero et al. 2015). Neutrinos on the other hand, are detected with the use of large volume underground water or ice tanks that are able to capture the Cherenkov light produced by the products of neutrino interactions in the medium. The IceCube neutrino observatory (Aartsen et al. 2017) at the South Pole is one of the largest active neutrino observatories with thousands of sensitive photomultiplier tubes distributed over a cubic kilometre of ice. Finally, cosmic rays are mainly detected by space-based instruments such as PAMELA (Menn et al. 2013) and the Alpha Magnetic Spectrometer on the International Space Station (Aguilar et al. 2014), which mainly use magnetic spectrometers to measure the charge and the sign of particles at certain energies.

What makes the observation of dark matter via these methods *indirect* is the fact that the targets of the observation – i.e. the dark matter particles – do not interact directly with our measuring instruments, rather, it is only the products of these particles that come into direct contact with our detectors causing the detectable signal. It should also be noted that although only this type of searches is labelled as 'indirect' by the physics community, based on the terminology adopted in Section 5.1, the observations of dark matter via its gravitational effects and cosmological precision measurements are also highly indirect in that the entities interacting with our measuring devices are not dark matter particles, rather they are other physical entities that are believed to be causally related to dark matter. For instance, in the observation of dark matter via its gravitational effects, what is eventually interacting with our telescopes is the light from the large scale structures that are causally related via gravity to the dark matter halos surrounding galaxies. This indicates that the indirectness of an observational method comes in various degrees depending on how long the causal series of events from the target (dark matter) all the way to the interaction of a physical entity with our measuring device is. In a sense, one might therefore say that the observation of dark matter via its gravitational effects and cosmological precision measurements is more indirect than the indirect searches of dark matter, however, this does not entail that these methods are either less informative or less reliable.

Collider searches: One of the most successful theories for the formation of dark matter in the early Universe is the freeze-out scenario.[15] A crucial assumption in the freeze-out mechanism is that the interactions between dark matter particles and standard model particles are sufficiently strong in order for the dark matter particles to enter into a thermal equilibrium with standard model states. If this is the case however, one would expect that the inverse process also has a sizeable cross section, and hence, particle colliders can be used to invert the process of dark matter annihilation in the early Universe. This is the main idea behind collider searches of dark matter at the Large Hadron Collider (LHC) at CERN: high-energy collisions of standard model particles are used for the direct pair-production of dark matter particles. Moreover, if we are lucky enough and the dark matter particle mass is comparable to the electroweak scale, LHC is also expected to be able to produce large quantities of dark matter particles via the decays of heavier states that are instantaneously created in high-energy proton-proton collisions. The bottom line is that collider searches are based on the possible interactions of dark matter particles with standard model quarks and gluons which, in either case, will lead to a missing energy (or missing momentum) signature in the final states due to the lack of interaction between dark matter and the material of the detectors, as well as to the detection of unexpected particle products.

There are two primary mechanisms by which the LHC could produce dark matter together with hadronic jets. According to the first mechanism, two strongly interacting dark matter particles are produced, and each one subsequently decays into dark matter and standard model particles, resulting in missing momentum plus two or more jets of hadrons. The second mechanism takes place when both dark matter and additional radiation are directly produced from the proton-proton collision, resulting in missing momentum recoiling against a 'mono-jet'. These two mechanisms give rise to five main processes in which dark matter pair-production could occur at the LHC: (i) mono-jet (ii) mono-V (iii) mono-Higgs (iv) dark matter with top quarks and (v) invisible Higgs decays.[16] In mono-jet processes the dark matter particles are produced in association with one or more QCD jets, in mono-V they are produced in association with a vector boson and so on.

---

[15] In the early Universe, highly energetic particles were created and existed in thermodynamic equilibrium – i.e. in a state where heavier particles were constantly converted to lighter ones and vice versa via mechanisms such as pair production, annihilation, interaction with other particles etc. As the universe expanded, the density of a particular particle species became too low to support frequent interactions and maintain the conditions for thermal equilibrium. At this point particles are said to '*freeze out*' (or decouple) and their number density, which is no longer affected by interactions, remains constant. The density of a particle species at the time of 'freeze-out' is known as the relic density (or relic abundance).

[16] For a detailed review on LHC dark matter searches see Kahlhoefer (2017).

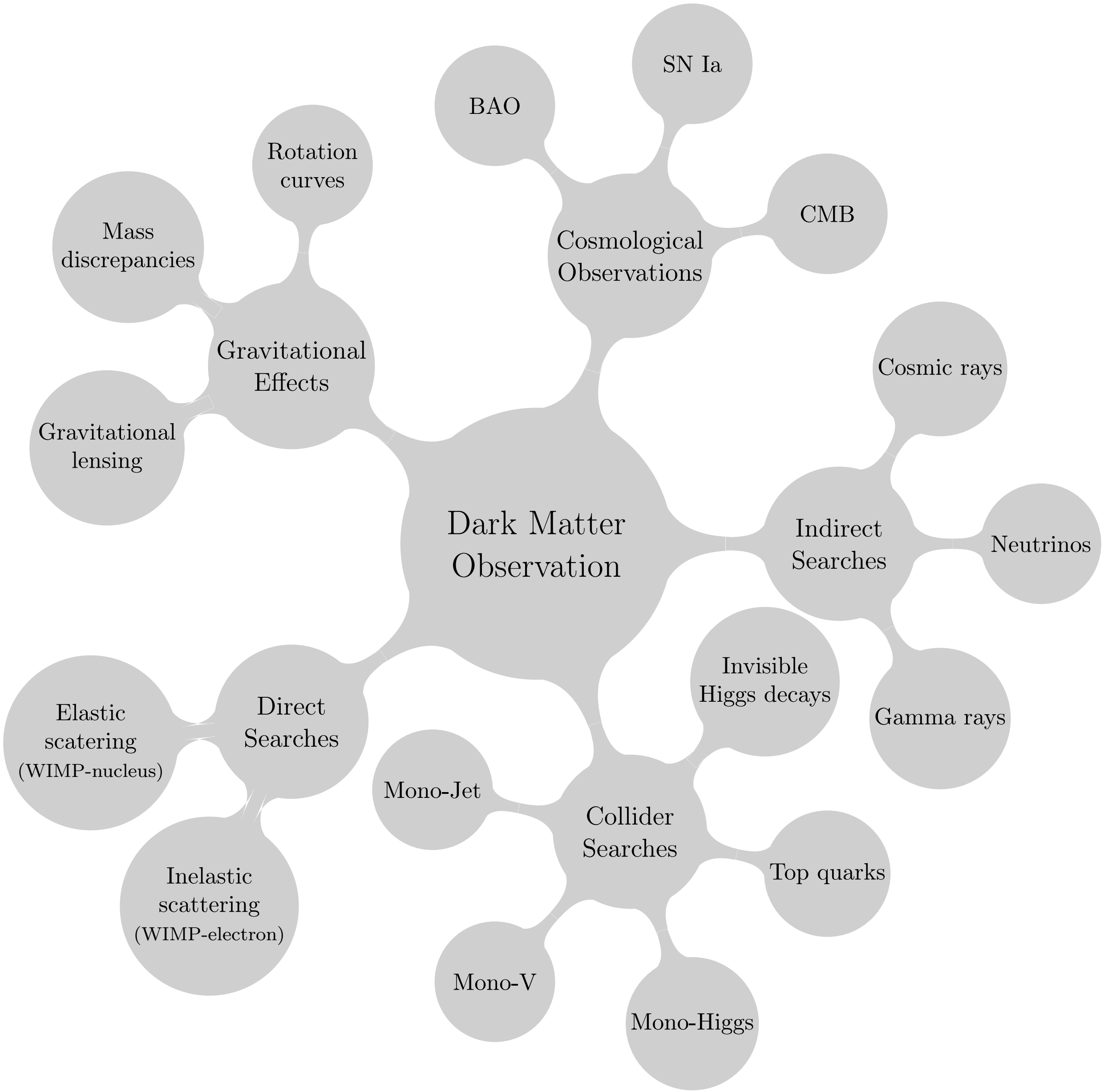

Figure 5.2: Dark matter observation: The first level of the diagram shows the five possible methods of dark matter observation (i) gravitational effects (ii) cosmological observations (iii) direct searches (iv) indirect searches and (v) direct searches. The second level of the diagram shows the various phenomena responsible for each type of observation.

Note finally that given that dark matter particles do not directly interact with the detectors of the LHC, strictly speaking, the observation of dark matter in these cases would be once again indirect, although physicists often talk of a direct process due to the potential direct pair production of dark matter particles from proton-proton collisions.

The five different methods for dark matter observation described above and the relevant underlying physical phenomena are summarised in Fig.5.2.[17] As already mentioned in the introduction, the fact that the available empirical evidence from the gravitational effects and the cosmological observations of dark matter place very little constraints on the properties of dark matter (recall Martens' very thin common core), has naturally led to the development of a diverse class of competing model scenarios of dark matter. The crucial question is which of these models best represent the nature and behaviour of dark matter and can eventually be used to derive further predictions on the cosmological scale. A possible way to tackle this question is to take advantage of the fact that dark matter can be probed by a variety of methods and deploy robustness arguments from the variability of experiments to support certain model hypotheses: if a specific model-hypothesis can be corroborated by a variety of experimental methods, then we have strong reasons to believe that the hypothesis under investigation is correct and the relevant model adequately captures the nature of dark matter.

The remaining of this chapter aims to show that, unfortunately, such arguments have a very limited scope in dark matter research since they can only be used for the comparison of different models that belong to the same model scenario for dark matter. To see why, it is useful to first examine some of the epistemic virtues of the various methods for dark matter observation in terms of their *informativeness*, *model sensitivity* and *reliability*.

5.3 Informativeness, Model Sensitivity, and Reliability

A necessary condition for employing robustness arguments from the variability of experiments in order to ensure the reliability of results from the various methods of dark matter observation, is that these methods are actually probing the same quantities. Siska de Baerdemaeker highlights this point in her discussion of the implications of methodological pluralism in dark matter research by noting that 'a crucial condition for measurement robustness is that the *same parameter or quantity* is being measured by the different

---

[17] A recent proposal for dark matter observation which is not included in our discussion is via the 'billion tiny pendulums' experiment (Carney et al., 2020). This is a novel approach in which dark matter can be directly observed based on its gravitational effects (instead of the recoil energy of nuclei) and hence, one might say that it is a special case in which two of the above methods are combined (direct searches and gravitational effects). The main idea of the experiment is that an array of quantum-limited mechanical impulse sensors (acting as a billion tiny pendulums) may be capable of detecting the correlated gravitational force created by a passing dark matter particle. However, at the time of writing, this is still a preliminary proposal and its merits cannot be fully evaluated.

experiments' (2021, p.140, emphasis added). She then suggests that the common core within the different experiments is provided by the fixed definition of the target of the observation. In her own words, 'the definition of the target system remains fixed under the employment of different methods. It provides, a common core that might underlie multiple methods attempting to probe the same target. Without this agreement on the common core, it is not obvious that methods that detect different phenomena are still probing the same target system and that measurement robustness arguments therefore apply' (*ibid.*).

While it is true that all five methods discussed above share the common goal of probing the physics of dark matter and the results they provide do indeed concern the properties of dark matter as de Beardemaeker notes, what I aim to show is that the thin definition of dark matter alone does not suffice to ensure that the agreement of results between different methods can be achieved. In order to establish that the results of two different methods are reliable via robustness arguments, we first need to ensure that the extracted information concerns the properties of the same dark matter models and relies on the same assumptions. However, as will be shown, this is rarely the case in dark matter research. The concepts of *informativeness* and *model sensitivity* will help us clarify this point.

Informativeness. The informativeness of a method concerns its ability to provide information on a number of different properties of dark matter either by providing specific values for these properties or – as it is often the case in dark matter research – by constraining the parameter space of a model. In general, a method E_1 is more informative than a method E_2 if the former is able to determine or constrain a set of dark matter properties P_1 that is larger than the set of properties P_2 determined by the latter. This rather simple definition of informativeness is all well and good in theory, but unfortunately, it only goes so far in the real world of experimental physics. In practice, the nature and the amount of information about the properties of dark matter that can be inferred from an observational method is determined by a number of various factors which make the comparison of different methods a much more complicated – if not impossible – process.

To begin with, most of the time the various methods of dark matter observation provide information about fundamentally different properties of dark matter. For instance, cosmological observations of dark matter based on CMB measurements are the sole source of information about the cosmic relic density of dark matter, since none of the remaining methods is able to provide any kind of information regarding the present quantity of dark matter in the universe. However, these observations provide no clue whatsoever about the possible non-gravitational interactions between dark matter and baryonic matter. The current constraints on the possible cross sections in the interactions between dark matter and standard model particles can only be inferred on the basis of direct, indirect and collider searches for dark matter, but not from the remaining two methods that are based on gravitational interactions and cosmological measurements. Hence, although the informativeness of a single method can be assessed by counting the number of dark matter properties for which it yields information, a straightforward comparison of the informativeness of different methods is often not possible due to the simple fact that these methods provide information about different properties of dark matter and are thus complementing each other.

A more important complication however, is that even in the cases where the various methods are yielding the same information (e.g. constraints on the mass of the dark matter particle), this information is conditional on a number of factors which vary in each experiment. In order to extrapolate a meaningful result from an experimental process, numerous assumptions need to be implemented both in the construction of the experimental apparatus and in the analysis of data. Ensuring that two different methods are probing the same quantity/parameter and provide concordant results thus requires taking into consideration the effects of these assumptions in the results of the experiment.

In the case of dark matter observation these factors can be grouped into three different categories: (i) the experimental models of the experiment (ii) the extrapolating assumptions and (iii) the model scenarios of dark matter and their sub-models.[18]

As we have seen in the previous chapter (Section 4.1), experimental models cover a broad category of models referring to every possible modelling activity that facilitates the construction of an experiment and the completion of a measurement process. For instance, in collider experiments this includes the various competing physical models for calculating the interactions of the produced particles with the different parts of the detector, while in direct searches such models would describe the ionization process of the liquid detectors and the interactions of photons with photomultiplier tubes. Experimental models also cover the required modelling for the simulation of large-scale dark matter

---

[18] It should be noted that these three sets of assumptions are not always entirely independent of each other. Rather, the above categorisation is a useful conceptual tool in order to illustrate how the extraction of information from an observational method is conditional on a number of assumptions which differ in scope.

formation that is often necessary for making measurements of dark matter properties.[19] As expected, the interpretation of data from an experiment and the further extrapolation of results strongly depends on the adopted experimental models in the various stages of the experiment, since the implementation of different models would give rise to a different set of data. The final effects of the adopted experimental models are often implemented in the results in the form of uncertainties, although it is also possible that a number of different results is derived, based on the selection of a specific combination of models.

Extrapolating assumptions are those assumptions needed for carrying out the required calculations for deriving information about the properties of dark matter, *after the acquirement of data from an experiment.* A profound example of a set of extrapolating assumptions comes from the interpretation of experimental results in direct searches for dark matter. The results from direct searches are necessarily extrapolated on the basis of some standard simplified assumptions such as the local density ρ_0 of dark matter, an isothermal profile of dark matter density and a Maxwell-Boltzmann velocity distribution. In the absence of knowledge about the exact properties of dark matter in the local region, the introduction of these assumptions is essential for carrying out the necessary calculations for the derivation of constraints on various properties of dark matter.

Finally, the information an observational method yields, also depends on the model scenario under consideration and its various sub-models. For instance, indirect searches are based on the fundamental assumption that dark matter annihilation and decay produces Standard Model particles in the form of gamma rays, neutrinos and cosmic rays. Similarly, collider experiments are only able to provide constraints based on the assumption that dark matter consists of WIMPs that can be produced in high-energy collisions. However, as already mentioned, the model scenario for WIMPs covers a broad class of specific models and depending on which sub-model is taken into consideration, a method can produce more than one set of results. That is, depending on which sub-model provides the assumptions needed for the interpretation of data, the experimental data from an observational method often provide different sets of constraints for different models of dark matter.

Model Sensitivity. The fact that the extrapolation of results from a particular method crucially depends on model related assumptions implies that the extracted information from a particular method of observation is most of the time *model specific.* For instance,

---

[19] See Gueguen (2020) and Smeenk & Gallagher (2020) for philosophical discussions on the uses of simulation in cosmology.

	Gravitational	Cosmological	Direct	Indirect	Collider
Collisionless DM	✓				
SIDM	✓			✓	
WIMPs		✓	✓	✓	✓
Sterile Neutrinos				✓	
Axions			✓		
Hidden/Complex DM			✓	✓	✓
Light Gravitinos					✓

Table 5.1: Model sensitivity of the various methods of dark matter observation

mass constraints on dark matter particles from collider experiments rely on the assumption that dark matter particles are WIMPs that interact weakly with baryonic matter, since if dark matter consists from purely collisionless particles such constraints cannot be derived from particle collisions. Model sensitivity is the epistemic virtue that concerns the ability of a method to provide information on a range of different model scenarios of dark matter and their sub-models. In general, a method E_1 is more model sensitive than a method E_2 if the former is able to determine or constrain, in one way or another, a set of models M_1 that is larger than the set of models M_2 determined by the latter.

Arguably, a good example of a highly model sensitive method comes from the indirect searches of dark matter, since these experiments are able to provide constraints on a number of different model scenarios including self-interacting dark matter, WIMPs, sterile neutrinos and models of complex dark matter. On the other hand, the cosmological observation of dark matter via precision measurements of CMB observables is considered to be highly model insensitive since the constraints derived from these methods cannot, in general, be applied to specific models. For instance, measurements of temperature anisotropies on the CMB provide the total relic density and the stability of dark matter on the cosmological scale, but insofar as it is possible that dark matter is made from more than one components, this information places no model-specific constraints in the relevant models for the particle nature of dark matter. It is only under the assumptions that dark matter consists of a weakly interacting massive particle, that CMB measurements are able to provide constraints on the dark matter-proton scattering section (Glucevic et al. 2019), as well as on the mass and the dark matter annihilation rates of WIMPs (Natarajan 2013).

The upshot is that when taking into consideration the informativeness of a particular method of dark matter observation it is important to highlight the degree to which the constraints imposed by its results are tied to specific models. Table 5.1 illustrates a tentative depiction of the model sensitivity of each method of observation with respect to various model scenarios of dark matter at the present time. The tick indicates that a method can provide information for at least one parameter of the relevant model, but it should be noted that the situation might well change in the future. For instance, axion-like particles are expected to be searched for in next generation colliders (Bauer 2019) and there is also a possibility of directly detecting collisionless dark matter via its gravitational effects. The table also highlights the fact that WIMPs can be probed by all three methods of dark matter particle detection (direct, indirect, collider) as well as with cosmological observations, which partly explains their increasing popularity.

Reliability. The consideration of the informativeness and model sensitivity of each method provides a good way of evaluating the nature of information that can be extracted and the various models that can be constrained, corroborated or rejected by each method. What remains to be seen is how these two virtues are related to the reliability of each method, since ultimately, what is of utmost importance for achieving the necessary progress in dark matter research is whether the extracted information from a method is reliable and can be used to enrich the common core concept of dark matter. Roughly speaking, the reliability of a method concerns its ability to consistently provide robust and infallible results that accurately describe the physical world. A perfectly reliable method would thus be one whose results consistently have zero deviation from the actual values of the physical properties of dark matter.

Franklin (1989, Ch.6) has nicely summarised a number of epistemological strategies that are commonly used in physics experiments in order to justify the reliability of the experimental results. Since there is no Archimedean point from which one can determine whether a method is perfectly reliable or not, the implementation of these strategies is often used as a rational argument to *increase the confidence of the scientific community* that the method is reliable and its results should be used for the construction of further more accurate models. The two strategies that are particularly useful for our discussion are the *repetition of experiments*, and the *variability of experiments*. Intuitively, the most straightforward way to test the reliability of an observational method is to repeat the experiment multiple times, and, if possible, at different locations involving different research groups and different tokens of the same equipment. If the results of the same experiment conducted at different times and in different locations agree, then this is a good sign that the process is at least consistently carried out in an appropriate manner. This argument from the *repetition of experiments* has indeed been used to test the reliability of the results from the DAMA/LIBRA collaboration claiming evidence of dark matter particles in the galactic halo (Bernabei et al. 2008). One of the main reasons why the DAMA/LIBRA results remain controversial amongst the scientific community is the fact that several subsequent experiments such as Xenon100, CDMSII and CoGent have not succeded in reproducing the same results in their attempts.[20]

However, even if the subsequent experiments had produced the same results, there would still be a probability that these common results are misleading due to a common external error-inducing factor such as the penetration of helium into the photomultiplier tubes of the experiments (Ferenc et al. 2019). In order to eliminate this probability, scientists often resort to additional arguments for the reliability of results based on the variability of experiments. The main idea is that a possible agreement between the results of two different types of methods, automatically increases our confidence in their reliability since 'it would be a preposterous coincidence if the same patterns were produced by two totally different kinds of physical systems'.[21]

Robustness arguments from the variability of experiments are thus inferences to the best explanation about the agreement of results obtained by two or more different experimental methods. The idea is that the best explanation for the fact that two distinct and different types of experiments provide agreeing results is that the common underlying theory explaining these results is correct. Such arguments have been widely used by philosophers in discussions about realism (e.g. Cartwright 1983; Salmon 1984; Massimi 2007), as well as in the context of measurement robustness (e.g. Woodward 2006; Stegenga 2009; Basso 2017).[22] The next and final section focuses on how such arguments can be used to increase the reliability of the experimental results in the observation of dark matter taking into consideration the informativeness and model sensitivity of each method. It will be argued that robustness arguments from the variability of experiments in dark matter research are only available within methods that are based on the same model scenario (e.g. WIMP searches via collider experiments and direct searches) and therefore face important limitations.

---

[20] For a nice discussion of the reliability and the robustness of the DAMA/LIBRA results and their controversy see Hudson (2009).

[21] Hacking (1983), as quoted by Franklin (1989, p.166). See also Franklin and Howson (1984) for a Bayesian argument showing how the variability of experiments increases the confidence of the scientists to the results in a higher degree compared to the repetition of the same experiment.

[22] Massimi's (2007, p.245) argument for the reality of unobservable phenomena on the basis of the variability of experiments is a characteristic example. In her own words: 'We are justified to believe in these [unobservable] phenomena [...] because they appear as stable and robust features (i.e. features that cannot be ascribed to background noise or to experimental error) detected through a variety of experimental procedures involving different kinds of data.'

5.4 The limits of robustness arguments

As already mentioned, in order to employ arguments from the variability of experiments to ensure the reliability of results in dark matter experiments, the results must concern the same parameter of the same model. The first and most straightforward complication in establishing this necessary condition comes from the fact that compared to the huge variety of models, there is relatively little overlap between different methods that are able to probe the same models, let alone the same parameters of these models. To put it simply, for a large number of viable candidate models for dark matter, obtaining results from more than one method is just not possible. This is the case, for instance, with axions and sterile neutrinos since, as Table 5.1 shows, the former can only be probed via direct searches, and the latter can only be probed via indirect searches. Moreover, even in the overlapping cases where some models can be probed by two or more methods, most of the time the extracted information concerns fairly small and disjoint areas of the relevant parameter space, making the comparison of results impossible. This is the case with WIMPs, for instance, which can be better probed in the low mass regions with colliders, whereas models with heavier particles are better constrained by direct searches. This complication reflects Stegenga's (2009) observations on the limits of robustness arguments based on the fact that scientists do not always have multiple techniques to generate common results, and that, often, the results obtained by multiple techniques are inconsistent or incongruent.

An additional complication however, concerns the fact that even in cases where multiple techniques are ostensibly able to provide concordant results, ensuring that these results are indeed in agreement is a very difficult – if not impossible – task. The main source of this difficulty relates to the informativeness and model sensitivity of each method and concerns the fact that the extracted information from the various different methods is always conditional on the experimental models, the extrapolating assumptions and the model dependence of each method. As we have seen, in order to extrapolate a meaningful result from an experimental process, numerous assumptions need to be implemented during the construction of the experimental setup and the extrapolation and analysis of data. Ensuring that two different methods are probing the same quantity/parameter and provide concordant results thus requires taking into consideration the effects of these assumptions in the results of the experiment.

The following remarks from Goodman et al. (2010) offer a rather illuminating example of how the results of an experiment are conditional to a number of factors. In a paper presenting a set of constraints on dark matter properties from collider experiments, the authors begin their discussion by stating that the interpretation of the results depends on the nature – and hence the adopted sub-model – of the dark matter particle:

> We consider the cases where the DM particle is a scalar [boson] or a fermion; if a scalar, it can be real or complex, and if a fermion, it can be Majorana or Dirac. Each of these cases is considered separately (*ibid.*, p.2).

They then continue by listing the extrapolating assumptions in the experiment in order to yield their results:

> We shall be considering the situation where the WIMP (which we will generically denote χ) is the only particle in addition to the standard model fields accessible to colliders. We will assume that χ is odd under some Z2 symmetry (e.g. R-parity in supersymmetry, or Kaluza- Klein parity in extra dimensions), and hence each coupling involves an even number of WIMPs with the lowest dimensional operators we consider containing two WIMPs. We assume whatever particles mediate interactions between the WIMPs and the SM fields are somewhat heavier than the WIMPs themselves, with their leading effect manifest as higher dimensional operators in the effective field theory. For simplicity, we assume the WIMP is a singlet under the SM gauge groups, and thus possesses no tree-level couplings to the electroweak gauge bosons. We also neglect couplings with Higgs bosons. (*ibid.*, p.2)

After the presentation of their results Goodman et al. proceed to conclude that the presented constraints on the strength of interactions of WIMPs with hadrons also depend on the mass of the dark matter candidate, as well as the coupling preference of dark matter: if dark matter primarily couples to gluons, the constrains from colliders become significantly tighter (*ibid.*,p.8).

The above remarks by Goodman et al. indicate that the various constraints placed on the interactions between WIMPs and hadrons from collider experiments are conditional on a set of introduced assumptions and are of course model specific to the model scenario of dark matter as weakly interacting massive particles. Given that different methods necessarily involve different assumptions, any comparison between these constraints and the constraints obtained by a different method (e.g. from direct searches) must therefore be made by taking into consideration the effects of these assumptions on the extrapolation of the constraints. Ultimately, what will establish a possible theory of dark matter is not a conclusive result derived by a specific method of observation, but rather, the combination of the results derived by these inherently different techniques. This is the aim of compatibility and complementarity studies in dark matter research, however, the severe lack of such studies highlighted by many physicists (e.g. Bauer et al. 2015), indicates the degree of difficulty for achieving this task. The current situation in dark matter research comprises a vast collection of largely unrelated papers placing *model specific constraints* on different model scenarios of dark matter, without examining the possible concordance of their results with alternative experiments. And while there is some hope that the results from different methods concerning a particular model scenario can be cross-checked for their reliability, the situation remains obscure for those results that are limited to a particular method.

A possible strategy to mitigate the model sensitivity of experiments is by conducting *model independent searches*. Roughly speaking, model independent searches aim to remove as many model specific assumptions as possible from an experiment by reducing the total number of assumptions within a class of sub-models, to a minimum set of common assumptions shared by all models. This strategy is quite similar with what is often described in philosophy of science as the *robustness analysis of models* (Weisberg 2006; Lloyd 2010) that can be employed in cases where various competing models are equally supported by a body of evidence, as is the case with dark matter. The aim is to locate and isolate a common set of assumptions in these models which lead to a shared prediction that does not depend on any specific assumptions of a particular model. In dark matter research, this practice is captured by the employment of simplified models of dark matter, briefly discussed in Section 5.2, however, as we have seen the range of these models is once again restricted to the sub-models of a specific model scenario (mainly WIMPs) and does not involve the extrapolation of common predictions from different model scenarios.[23] In other words, model independent searches based on simplified models are typically independent from the various sub-models of a particular model scenario, but they are not entirely independent of every model scenario of dark matter. Model independent measurements in a strong sense, such as the calculation of the cosmic abundance via measurements on the CMB are possible, but as we have already seen they are at the same time model insensitive and hence provide little information as to which of the competing models best represents the physical world.

---

[23] See Morgante (2018) for a review of simplified models in dark matter.

5.5 Conclusions

To sum up, we have seen that the minimum set of constraints from the thin common core concept of dark matter allows the development of a relatively large and diverse number of model scenarios for the nature of dark matter, that are, nonetheless, equally viable. The range of these models nicely illustrates how phenomenological models in physics are often developed to provide how-possibly explanations of various unexpected phenomena and data. Our brief discussion of simplified models of dark matter in which different models with common assumptions are grouped together (e.g as WIMPs) in order to derive their common predictions, illustrates yet another nice example of how Mach's principle of the economy of thought is ubiquitous in the methodology of physics. Our main aim however, has been to investigate whether the abundance of viable models for the nature of dark matter and the fact that there are various methods available for probing the properties of dark matter can be used to construct robustness arguments from the variability of experiments.

The conclusion is that given the model dependence of experiments and the introduction of various assumptions for the extrapolation of information in dark matter observation, robustness arguments from the variability of experiments in dark matter research are limited to the results concerning the sub-models of a specific model scenario. The first implication of this limitation is that robustness arguments can only be used for models that can be probed via more than one method, such as WIMPs and self-interacting dark matter, but are unavailable for model scenarios of dark matter which depend on a single method, such as axions and sterile neutrinos. The second and most important implication is that robustness arguments from the variability of experiments can only be used for the comparison of the various competing sub-models of a specific model scenario such as the various WIMP sub-models. The problem is that while this is a good practice to distinguish between competing sub-models of a specific dark matter model scenario, it leaves the essential question about the nature of dark matter untouched. No matter how well corroborated a specific sub-model is, the fact that the data are model dependent in the first place means that there is no way of ensuring that any alternative scenarios for the nature of dark matter should be neglected.

This situation is characteristic of a broader problem in physics concerning the theory ladenness of observation that we have already discussed in the previous chapter. In order to build experiments and interpret the data from them, numerous assumptions about the nature of the involved entities need to be implemented. In dark matter research, these assumptions are often related to the particle nature of dark matter and its properties concerning its self annihilation and interaction with baryonic matter. The existing constraints on the mass of the dark matter particle, the cross sections of its interactions with ordinary matter, the self annihilation rates etc. are all extrapolated based on the assumption that dark matter is described by a particular model scenario (i.e. constraints are model specific). This helps in constraining the parameter space of various models in that it tells us what the nature of dark matter should be *if* it is captured by a specific model scenario. The crucial question however, is which model(s) among the viable candidates corresponds to the actual nature of dark matter. To answer this question, the thin common core concept of dark matter needs to be enriched in order to restrict the range of viable models. However, as we have seen, the only possible way to enrich the common core by adding further constraints is by presupposing the very same models that we are trying to constrain, and for which there is no independent observational evidence at the first place.

This fact gives rise to a puzzling situation regarding progress in dark matter research. As understood within the context of this chapter, progress in the science of dark matter amounts to narrowing down the number of viable dark matter models by enriching the common set of constraints from which these models are built. In order to enrich this common set of constraints, more reliable results have to be obtained via the available methods of dark matter observation. Our discussion has highlighted the difficulty of this task and the limitations of robustness arguments in ensuring the reliability of the extrapolated results. On the one hand, robustness arguments are necessarily confined within results concerning the sub-models of a particular model scenario, and hence provide no information about the viability of alternative competing model scenarios. On the other hand, model independent results can enrich the common set of constraints, however, by definition, they place no specific requirements on the various competing models of dark matter. This leaves us with a puzzling situation: how are we to enrich the common core concept in order to suppress the underdetermination of viable models? The solution to this puzzle, as well as to the aforementioned problems with robustness arguments, amounts to figuring out a possible methodology for extrapolating model independent results which, at the same time, are able to restrict some model scenarios and preclude others. The present chapter illustrates how scientists and philosophers of science can work together towards the identification and rectification of these matters.

Chapter 6

Conclusions and Open Questions

6.1 Key conclusions

This thesis investigates the methodology of modern physics, with an emphasis on the scientific practice of modelling. In the first two chapters, a more generic approach was taken with a critical discussion of the existing literature and an outline of the conceptual foundations of the thesis (Chapter 1), followed by an intervention on the current literature on the ontology of models (Chapter 2). In the remaining three chapters, the focus was shifted to more practical issues and challenges in the methodology of modern physics and the three main axes of the thesis were elaborated in each one of these chapters. The first axis concerned the relationship between models and background theories (Chapter 3), the second concerned the relationship between theoretical models and experimental data (Chapter 4), and the third concerned the observation of unobservable entities in physics, such as dark matter particles (Chapter 5).

The study of the various aspects of the methodology of modern physics across the different fields of theoretical particle physics, experimental particle physics and cosmology highlights the various functions of models in serving the most important aims of physics as a scientific discipline. It also highlights the different ways in which models are built via the manipulation of the fundamental principles of a theory and the use of controllable or non-controllable idealizations, as well as via the incorporation of ad hoc hypotheses in order to fit the experimental data. Surely, if we wish to understand science and its methodology, a detailed study of the structure of the different types of models and their roles is essential. This thesis aims in making a substantial contribution towards this end by engaging in a descriptive and normative analysis of the science and the philosophy of physics.

If there is a one-sentence general conclusion to be drawn from this thesis, it is that the art of modelling is a multifaceted practice within the methodology of modern physics, and to the extent that the success of science can be measured by the achievement of accurate predictions, good explanations and ways of economising human thought, the construction of different types of models is an indispensable part of it. Scientific models are multifunctional epistemological tools of investigation that come in various forms and serve different purposes, and as such, they play an essential role towards the achievement of the most important aims in physics. The ultimate aim of this thesis was to provide an in-depth analysis of how different types of models play different roles in science, demonstrating this way how philosophy of science can make fruitful contributions to science by providing the conceptual grounds for further scientific enquiry and clarifying with its rich analytical tools the various scientific challenges that are often found in the methodology of modern physics.

Perhaps the most significant specific conclusion of this thesis concerns the identification of an under-studied type of models, that is primarily found in perturbative quantum field theory, discussed in Chapter 3. While most of the relevant discussions in the existing literature are focused on the distinction between theoretical models and phenomenological models, our analysis in this chapter led to the identification of a third special type of models in physics whose relationship with the background theory cannot be described in clear and deductive terms. As we have seen, these models are neither theoretical, nor phenomenological, but nonetheless play a central role in one of the most successful areas of modern physics, and as such, they deserve to be studied in more detail. The presence of these non-theoretical models in quantum field theory poses a challenge to the semantic view of scientific theories and to van Fraasen's commitment on theoretical models, but most importantly, it forces us to reconsider the relationship between theories and models in physics in terms of their empirical adequacy.

In the following chapter, our aim was to provide various insights on the nature and the role of data models in high-energy physics, not by merely providing a clear definition of data models, but rather, by highlighting the long and reiterative process for the construction of a data model in HEP and the importance of considerations regarding the selection criteria, efficiency calculations, data fitting, and the calculations of uncertainties. Our main conclusion in this chapter was that the final data model which is compared to the theoretical hypothesis is essentially a co-production of real and simulated data, a fact that is often neglected in the discussions about whether the former are epistemically superior than the latter. Most importantly, the close analysis of the actual scientific practice at one of the most important modern experiments in physics at CERN, revealed that although philosophy of science can make substantial contributions in framing the challenges and subtleties of scientific methodology, the construction of general (and formalized) philosophical frameworks describing the scientific practice is often a very difficult task, and always runs the danger of failing to capture the intricacies and the details of a specific research programme.

Finally, the study of the current situation in dark matter research in Chapter 5 whereby a number of different models about the nature of dark matter is equally supported by the available observations, illustrates the ways in which the analytical tools of philosophy of science can provide a potential explanation about the presence or absence of progress in physics. By drawing on the rich literature on robustness in philosophy of science, the aim in this chapter was to provide a useful epistemological basis for the evaluation of different methods of dark matter observation in terms of their informativeness, model sensitivity and reliability. The main conclusion was that a partial explanation about the slow progress in dark matter research is due to the limits of robustness arguments from the variability of experiments in enriching the common core concept of dark matter and eventually narrowing down the space of possibilities for the different model scenarios about the nature of dark matter.

In addition to the specific conclusions of each chapter, the study of the current scientific practice in this thesis nicely outlines an overall framework of the methodology of modern physics and the different roles of models in it. Figure 6.1 is an ambitious attempt to illustrate in a simple diagram the central components of the otherwise complicated scientific practice of modern physics, as they were highlighted throughout the thesis. At the top left corner, are the theories of instrumentation. These are the auxiliary physical theories that are typically used for the construction of the various instruments, such as telescopes and particle detectors, and the design of experiments for the observation of specific physical phenomena. Auxiliary theories of instrumentation can be anything from optics and condense matter physics, to quantum chemistry and fundamental theories of physics such as quantum field theory and general relativity. The role of these theories is to provide experimental models whose primary functions are, as we have seen in Chapter 4, to provide an experimental analogue of the primary theory to be tested and to specify the necessary techniques for linking the experimental data to the theoretical hypotheses

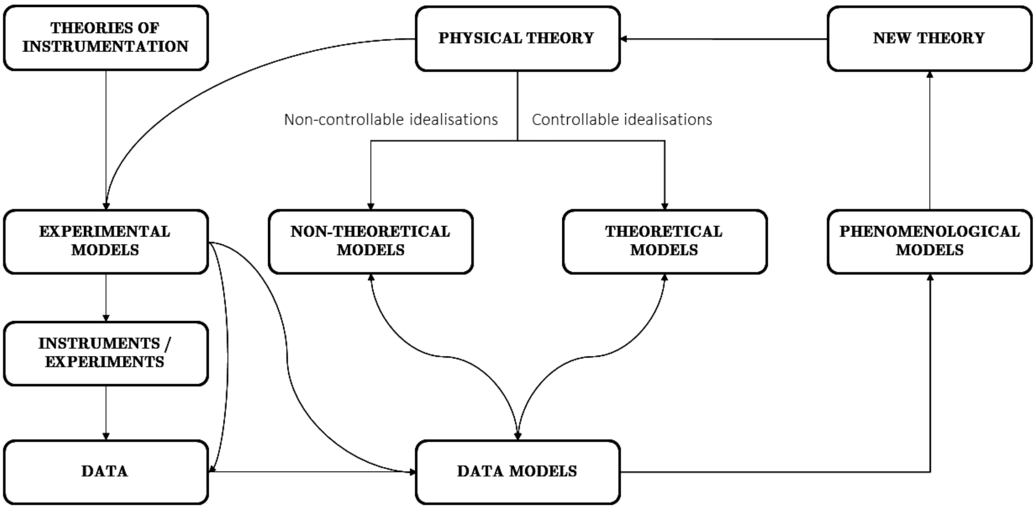

Figure 6.1: The methodology of modern physics.

under investigation. Some examples of experimental models are the physical models for calculating the interactions of particles with the different parts of the detectors in collision experiments, and models of nuclear physics for the interaction of dark matter particles with the nucleus of the atoms of noble gases in direct searches for dark matter. As the diagram shows, experimental models can be constructed both from auxiliary theories of instrumentation and the primary theory to be tested in an experiment – such as the standard model.

Once an experiment or an observational method is set up with the help of experimental models and the relevant theories, various techniques are then used for the extrapolation of raw experimental data that are then moulded into a data model. As illustrated by the diagram, both the extraction of the initial data from an experiment *and* the process of refining the data into a data model is facilitated by the use of various experimental models which include simulation modelling (event generators, detector simulations, pseudo-experiments with Monte Carlo simulations etc.), models of data acquisition, and statistical models for the analysis of data. Moreover, as made salient in Chapter 4, the journey from the initial data of an experiment to the final data model that is eventually compared with the theoretical hypothesis under investigation is a long and reiterative process of trial and error which includes the fusion of real and simulated data and the employment of highly sophisticated techniques for data analysis.

At the top centre of the diagram is the primary theoretical framework whose empirical consequences and predictions are put to the test by experiment. In Chapters 1 and 3, we discussed the ways in which the theoretical framework of the theory can be extended with the implementation of various mathematical techniques of approximation into a mathematical framework of a model. The employment of controllable idealisations gives rise to the theoretical models of the theory, whereas the use of non-controllable idealisations, such as in perturbative quantum field theory, provides non-theoretical models. The empirical consequences and predictions of the theory under investigation are produced by the mathematical manipulation of these models and are connected to the natural world via an appropriate interpretation. These empirical consequences – which we have also labelled as the theoretical hypotheses H_0 in Chapter 4 – are then compared to data models from experiments, given that the latter are expressed in a language that is compatible with the (mathematical) language of the theory. In case of agreement – such as in the experimental validation of the magnetic moment of the electron discussed in Chapter 1 – the background theory is corroborated and the collective degree of belief for the empirical adequacy of the theory is increased. In case of disagreement – such as in the case of the suspected violation of Lepton Flavour Universality at the LHCb (Chapter 4) and the presence of dark phenomena in cosmological observations (Chapter 5) – a number of new phenomenological models are developed with the primary aim of accommodating the deviating data.

In order to provide results that fit the new unexpected data, these phenomenological models often deliberately violate fundamental principles of the primary theoretical framework. Throughout the thesis we have seen several examples of such models: in Chapter 1 we introduced the concept of phenomenological models with the historical example of Bohr's model of the Hydrogen atom, in Chapter 4 we briefly discussed the emergence of models containing new types of particles to accommodate the possible violation of Lepton Flavour Universality, and finally, in Chapter 5 we have seen how the severe underdetermination of theory by the available data for the existence of dark matter gives rise to a host of viable phenomenological models of dark matter both in the microscopic and the macroscopic scales. The importance of phenomenological models in the methodology of physics lies in the fact that they are often the starting point for the refinement, or even the replacement of existing theories with a new theory. If a phenomenological model is particularly successful, its mathematical framework is then developed further in order to make new predictions and to accommodate, not only the initial deviating data from which it was constructed in the first place, but also other previously known results. The accommodation of previously known results by a phenomenological model and the confirmation of its predictions in a new domain often mark a significant milestone in physics whereby a well-established theory is refined, extended or replaced, in order to include the successful phenomenological model in its arsenal and explain all the relevant experimental data. The new theory then provides its new empirical consequences via new theoretical models that are put to the test by comparison with data models from experiments and so on.

In the special case of perturbative quantum filed theory where the supposedly theoretical models of the theory are produced via non-controllable idealisations (and hence they are not theoretical), there are two possible options. The first is to maintain the existing models and work on the mathematical justification of the approximating techniques, with the hope of making the uncontrollable idealisations controllable. This would be achieved in perturbative quantum field theory by (i) rigorously calculating the beta-functions in models of realistic quantum field theories and (ii) showing that the perturbative series in such models are Borel summable in order to justify strong asymptoticity. Unfortunately, the satisfaction of these two conditions remains for the time being a far fetched possibility. The second option is to see realistic quantum field theories, such as QED and QCD, as good approximations for the low energy regime of a future more complete theory whose theoretical models will accommodate the current experimental data. Nevertheless, the difficulty of rigorously deriving theoretical models from the theories of QED and QCD might be a sign that we are reaching the limits of our scientific knowledge in the particular domain of physics.

6.2 Revisiting the aims of physics

Going back to the discussion in the first chapter of this thesis, recall that we presented a framework for the characterisation of models in terms of four main features: (i) a mathematical framework, (ii) an interpretation, (iii) an ideal system, and (iv) the representational media. It was emphasised that this framework is not a definition of models, nor a list of necessary and sufficient conditions, but rather, a useful conceptual tool in order to better understand the nature of models in physics and facilitate the discussions in the following chapters. We can now see, how each of these components helps in understanding the various methodological and philosophical aspects discussed in this thesis.

First and foremost, the mathematical framework of the model is what determines the relationship of models with theories, with which we were mainly concerned in Chapter 3. The precise way in which the mathematical framework of a theory is enriched with mathematical approximations in order to provide the mathematical framework of the model determines whether the resulting model is a theoretical or a non-theoretical model. Moreover, we have also seen that one of the two defining characteristics of phenomenological models is the presence of conflicting mathematical sentences between the mathematical framework of the models and the mathematical framework of the theory. The second characteristic is the fact that phenomenological models are built to fit the experimental data that cannot be accommodated by the theoretical models of the existing theory.

This fit of models with data is established in virtue of the interpretation of the mathematical framework of the models. The crucial role of the interpretation of the mathematical framework is especially apparent in the case of experimental models, e.g. the physical models of the interactions between product particles and atoms of the detectors in collision experiments, whose main role is to capture the behaviour of the physical components of an experiment as accurately as possible. The interpretation of the mathematical framework of these models essentially establishes that certain mathematical relationships between variables of the model capture the relationships between measurable physical quantities of a physical system. One way to understand how the interpretation of a mathematical framework links the model with the physical world, is via the causal interpretation of the mathematical framework of a model, in which changes in dependent variables of the model that result from changes in the independent variables, correspond to changes in physical quantities as a result of the changes in other physical quantities with which they are causally connected.

As for the ideal system that often features in models and the various representational media by which models are expressed, these components were particularly helpful in the discussion of the literature on the ontology of models in Chapter 2. As we have seen, most of the metaphysical challenges in this debate come from the presence of abstract objects (or ‘missing systems’ in Thomson-Jones’ terminology) in models and the attribution of properties to them. Moreover, the various philosophical views that models *are* stylized descriptions of physical systems, mathematical equations, and physical objects indicate that the ontological status of scientific models has been sometimes identified with the representational media by which these models are expressed and communicated. However, this thesis has made salient that the concept of models in physics encloses much more than the representational media by which it is expressed, and hence, models are best viewed as multifunctional epistemological tools of investigation comprising various different elements.

In the first chapter of the thesis, it was also argued that models play a crucial role in achieving the three major outcomes of physics, namely the economic description of nature, the explanation of physical phenomena, and the prediction of novel experimental results. One of the aims of the model-based approach in the methodology of modern physics that was taken in this thesis was to illustrate this claim. Both prediction and explanation are based on the agreement of the theoretical hypotheses of models with the experimental results and observations of physical phenomena. The ultimate goal of modelling physical phenomena is to acquire *results* in the form of simple numerical answers, probability distributions, specific relations between two or more physical quantities etc., that will eventually be tested by experimental practices. As it was made clear in Chapter 4, the agreement of these results with the corresponding data models amounts to a prediction/accommodation of a new or a previously known result, whereas a disagreement provides hints for the failure of the existing theoretical models to explain certain phenomena and the need to refine existing theories.

The importance of phenomenological models in providing how-possibly explanations was particularly illustrated in Chapter 5 by showing how the presence of certain (dark) phenomena that cannot be accommodated by the present theoretical frameworks of general relativity and the standard model of particle physics gives rise to the development of a multitude of models of dark matter. Each of these models, whether at the micro-scale of particle physics or the macro-scale of cosmology, purports to explain the presence of such phenomena by providing a rigorous mathematical framework, and an appropriate interpretation that corresponds to the phenomenology of large scale structures in cosmology. The ultimate aim is to derive predictions from these models that will eventually be validated by experiments, and then embed these models in a broader theoretical framework that will eventually be able to produce theoretical models that will accommodate these dark phenomena. The example of the Lightest Supersymmetric Particle as a dark matter candidate is a prime example of how the present theoretical framework of the standard model can be extended in order to accommodate dark matter phenomena along with other known anomalies in the subatomic scale.

Finally, the contribution of models for the economy of thought was also emphasised in Chapter 5 by showing how the various models of dark matter can be classified and condensed into simplified models capturing the essential components of various different models for the particle nature of dark matter. By classifying various models of dark matter as models of Weakly Interactive Massive Particles (WIMPs) scientists are able to identify the common assumptions between these models and test the predictions that follow, without the need of developing and studying in detail each one of these models. Another way in which the economic classification and description of nature via modelling becomes apparent, is the construction of models for the classification of various astronomical objects such as neutron stars and primordial black holes as dark matter candidates. For instance, by constructing a model of a neutron star as a candidate for dark matter, scientists are essentially constructing a single mathematical framework describing the common behavioural patterns of a host of different physical systems that classify as neutron stars and could potentially be responsible for certain dark phenomena via their gravitational effects.

6.3 Future work

As a closing remark, let us highlight the most important open questions emerging from each chapter, and point out some possible directions of future work suggested by this thesis. Chapter 2 nicely illustrates how Carnap's pragmatic approach to metaphysics is still relevant to contemporary debates in the philosophy of science, such as the debate on the ontology of models. An interesting project is to see how and if this pragmatic approach can be applied not only to other contemporary debates in philosophy of science – e.g. to the discussions on scientific representation and natural kinds – but also to purely scientific debates regarding the existence of certain physical entities such as dark matter particles. For instance, could the statement about the existence of a collisionless dark matter particle that has no interactions with baryonic matter other than via the gravitational force be understood as an internal sentence within a useful linguistic – yet scientific – framework for the explanation of dark phenomena? And if yes, what are the consequences of adopting such a pragmatic stance in science?

In Chapter 3, the main conclusion was that the relationship between the realistic models of perturbative quantum field theory and the background theory cannot be explained in deductive terms. An open question that was briefly discussed in the end of the chapter was whether the empirical adequacy of the models can be attributed to the theory as well, given that the former are products of non-controllable idealisations. Further work is needed to understand the terms in which the empirical adequacy of the theory is established by the success of its models in such cases. Moreover, towards the end of the chapter we briefly mentioned that the lack of a rigorous mathematical justification of the introduced idealizations combined with the fact that there is no clear structural characterisation of a physical system in these models poses some interesting questions about the explanatory status of the non-theoretical models in perturbative QFT. This is yet another very interesting philosophical issue which deserves to be studied in more detail. Finally, another interesting issue came to the surface when discussing the justification of the behaviour of quantum field theories in high energies via the beta function (Sec. 3.3). As we have seen, the justification for the approximating techniques of renormalisation and regularisation is attempted by appealing to further approximations for the calculation of the beta function in high-energy regimes. This is an instance of a philosophically interesting situation in physics in which the justification of an approximation is achieved by introducing further approximations which presumably require further justification and so on. Future work is needed to identify more examples of this regressive practice in physics and determine the conditions under which such regress might or might not be vicious.

In Chapter 4, the detailed case study of the experimental practice at the LHCb revealed amongst other things the various pragmatic considerations during the process of data analysis in large-scale experiments – e.g. with respect to the choice of models for the interactions of product particles with the detector, the available computational time and power in the subsystems of the detector, the storage capacity at the CERN Data Centre, the reconstruction of events so that they are amenable to statistical analysis and so on. The exact nature of the impact of these pragmatic considerations on the final result is an interesting topic worth pursuing further. What is the dependence of the final result on pragmatic considerations during the experiment and which of these considerations have the biggest impact? How can the impact of pragmatics be mitigated? Moreover, the analysis in this chapter also revealed the important role of simulation in various aspects of the experiment, from the calculation of detector efficiencies, to the design and optimization of the detectors in the early stages of the experiment and the calculation of the performance of the detector, which are crucial for the extraction and interpretation of the available data. It also highlighted the fact that the final data model is essentially a co-production of real and simulated data that are blended in various stages of the experiment. These observations give rise to a number of interesting issues regarding the exact role of simulations and their impact on large-scale experiments in high-energy physics. Does the inclusion of simulated data in data models undermine in any way the pureness of the final result? Is there a way to identify and even quantify the ways in which simulation determines the nature of the final results? And does the necessity of the use of simulations in these experiments signal a new era of simulation-based science?

Finally, the discussion of dark matter research in the final chapter brought to the surface some of the most important challenges in the field, indicating possible ways for future collaborations between philosophers and physicists. The proliferation of viable dark matter candidates and the observation that a large part of current work on dark matter consists of largely unrelated and model-specific constraints on different model scenarios of dark matter, highlights the need for a common ground of reference – that could take the form of an online database – where these works can be compared and connected with each other. The design and construction of this reference source could be achieved via a collaboration of physicists and philosophers of science. Moreover, further work needs to be done to elucidate the major cause for the puzzling situation in dark matter research discussed in the end of the chapter, in order to identify the possible ways for making further progress in the search for dark matter. A collective and rigorous conceptual and methodological analysis of the scientific practice in dark matter research by philosophers of science could eventually shed light on the various methodological conundrums in this field and outline possible strategies for achieving further progress in dark matter research by enriching the common core of dark matter and reducing the number of viable models. The work presented in this chapter, is a first step towards this direction.

BIBLIOGRAPHY

Aartsen, M. G., M. Ackermann, J. Adams, J. A. Aguilar, M. Ahlers, M. Ahrens, D. Altmann, K. Andeen, T. Anderson, and I. Ansseau (2017). The IceCube Neutrino Observatory: instrumentation and online systems. *Journal of Instrumentation 12*(03), P03012.

Acero, F., M. Ackermann, M. Ajello, A. Albert, W. B. Atwood, M. Axelsson, L. Baldini, J. Ballet, G. Barbiellini, and D. Bastieri (2015). Fermi large area telescope third source catalog. *The Astrophysical Journal Supplement Series 218*(2), 23.

Achinstein, P. (1968). *Concepts of Science: A Philosophical Analysis.* Baltimore: MD: Johns Hopkins Press.

Aghanim, N., Y. Akrami, M. Ashdown, J. Aumont, C. Baccigalupi, M. Ballardini, A. J. Banday, R. B. Barreiro, N. Bartolo, and S. Basak (2018). Planck 2018 results. *VI. Cosmological parameters 1807.*

Aguilar, M., D. Aisa, A. Alvino, G. Ambrosi, K. Andeen, L. Arruda, N. Attig, P. Azzarello, A. Bachlechner, and F. Barao (2014). Electron and positron fluxes in primary cosmic rays measured with the Alpha Magnetic Spectrometer on the International Space Station. *Physical review letters 113*(12), 121102.

Akerib, D. S., X. Bai, S. Bedikian, E. Bernard, A. Bernstein, A. Bolozdynya, A. Bradley, D. Byram, S. B. Cahn, and C. Camp (2013). The large underground xenon (LUX) experiment. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 704*, 111–126.

Alspector-Kelly, M. (2001). On Quine on Carnap on ontology. *Philosophical Studies 102*(1), 93–122.

Ankeny, R., H. Chang, M. Boumans, and M. Boon (2011). Introduction: philosophy of science in practice. *European journal for philosophy of science 1*(3), 303.

Antoniou, A. (2021a). A pragmatic approach to the ontology of models. *Synthese 199*(3), 6645–6664.

Antoniou, A. (2021b). What is a data model? *European Journal for Philosophy of Science 11*(4), 1–33.

Aoyama, T., T. Kinoshita, and M. Nio (2018). Revised and improved value of the QED tenth-order electron anomalous magnetic moment. *Physical Review D 97*(3), 36001.

Archilli, F., W. Baldini, G. Bencivenni, N. Bondar, W. Bonivento, S. Cadeddu, P. Campana, A. Cardini, P. Ciambrone, and X. C. Vidal (2013). Performance of the muon identification at LHCb. *Journal of Instrumentation 8*(10), P10020.

Bailer-Jones, D. (2009). *Scientific models in philosophy of science.* Pittsburgh, PA: University of Pittsburgh Press.

Bailer-Jones, D. M. (2002). Models, metaphors and analogies. *The Blackwell guide to the philosophy of science 114*.

Bailer-Jones, D. M. (2003). When scientific models represent. *International studies in the philosophy of science 17*(1), 59–74.

Bailey, D. C. (2017). Not Normal: the uncertainties of scientific measurements. *Royal Society open science 4*(1), 160600.

Baker, A. (2003). Quantitative parsimony and explanatory power. *The British Journal for the Philosophy of Science 54*(2), 245–259.

Banks, E. C. (2004). The philosophical roots of Ernst Mach's economy of thought. *Synthese 139*(1), 23–53.

Barlow, R. (2002). Systematic errors: facts and fictions. *arXiv preprint hep-ex/0207026*.

Barnes, E. (2021). Prediction versus Accommodation. *The Stanford Encyclopedia of Philosophy*.

Basso, A. (2017). The appeal to robustness in measurement practice. *Studies in History and Philosophy of Science Part A 65*, 57–66.

Batterman, R. W. (2001). *The devil in the details: Asymptotic reasoning in explanation, reduction, and emergence.* Oxford University Press.

Batterman, R. W. (2009). Idealization and modeling. *Synthese 169*(3), 427–446.

Batterman, R. W. and C. C. Rice (2014). Minimal model explanations. *Philosophy of Science 81*(3), 349–376.

Bauer, D., J. Buckley, M. Cahill-Rowley, R. Cotta, A. Drlica-Wagner, J. L. Feng, S. Funk, J. Hewett, D. Hooper, and A. Ismail (2015). Dark matter in the coming decade: Complementary paths to discovery and beyond. *Physics of the Dark Universe 7*, 16–23.

Bauer, M., M. Heiles, M. Neubert, and A. Thamm (2019). Axion-like particles at future colliders. *The European Physical Journal C 79*(1), 74.

Beauchemin, P.-H. (2017). Autopsy of measurements with the ATLAS detector at the LHC. *Synthese 194*(2), 275–312.

Bečirević, D., S. Fajfer, N. Košnik, and O. Sumensari (2016). Leptoquark model to explain the B-physics anomalies, R_K and R_D. *Physical Review D 94*(11), 115021.

Bernabei, R., P. Belli, F. Cappella, R. Cerulli, C. J. Dai, A. D'Angelo, H. L. He, A. Incicchitti, H. H. Kuang, and J. M. Ma (2008). First results from DAMA/LIBRA and the combined results with DAMA/NaI. *The European Physical Journal C 56*(3), 333–355.

Bertone, G., D. Hooper, and J. Silk (2005). Particle dark matter: Evidence, candidates and constraints. *Physics Reports 405*(5-6), 279–390.

Bertone, G. and T. M. P. Tait (2018). A new era in the search for dark matter. *Nature 562*(7725), 51–56.

Bifani, S., S. Descotes-Genon, A. R. Vidal, and M. H. Schune (2018). Review of Lepton Universality tests in B decays. *Journal of Physics G: Nuclear and Particle Physics 46*(2), 23001.

Bird, A. and E. Tobin (2008). Natural Kinds. *The Stanford Encyclopedia of Philosophy*.

Bird, G. H. (1995). Carnap and Quine: Internal and external questions. *Erkenntnis 42*(1), 41–64.

Black, M. (2019). *Models and metaphors.* Cornell University Press.

Blackburn, S. (2006). Antirealist Expressivism and Quasi-Realism. In D. Copp (Ed.), *The Oxford Handbook of Ethical Theory.* Oxford: Oxford University Press.

Blake, T., G. Lanfranchi, and D. M. Straub (2017). Rare B decays as tests of the Standard Model. *Progress in Particle and Nuclear Physics 92*, 50–91.

Blatti, S. and S. Lapointe (Eds.) (2016). *Ontology after Carnap*. Oxford: Oxford University Press.

Blaum, K., Y. N. Novikov, and G. Werth (2010). Penning traps as a versatile tool for precise experiments in fundamental physics. *Contemporary Physics 51*(2), 149–175.

Bogen, J. and J. Woodward (1988). Saving the Phenomena. *The Philosophical Review 97*(3), 303–352.

Bokulich, A. (2011). How scientific models can explain. *Synthese 180*(1), 33–45.

Bokulich, A. (2014). How the tiger bush got its stripes: how possibly vs. how actually model explanations. *The Monist 97*(3), 321–338.

Bokulich, A. (2018). Using models to correct data: paleodiversity and the fossil record. *Synthese*, 1–22.

Bokulich, A. (2020). Towards a taxonomy of the model-ladenness of data. *Philosophy of Science 87*(5), 793–806.

Bokulich, A. and W. Parker (2021). Data models, representation and adequacy-for-purpose. *European Journal for Philosophy of Science 11*(1), 1–26.

Boon, M. and T. Knuuttila (2009). Models as epistemic tools in engineering sciences. In *Philosophy of technology and engineering sciences*, pp. 693–726. Elsevier.

Bordone, M., G. Isidori, and A. Pattori (2016). On the standard model predictions for R_K and R_{K^*}. *The European Physical Journal C 76*(8), 440.

Boumans, M. (1999). Built-In Justification. In M. Morgan and M. Morrison (Eds.), *Models as mediators : perspectives on natural and social sciences1*, pp. 66–96. Cambridge: Cambridge University Press.

Bovens, L., C. Hoefer, and S. Hartmann (2008). *Nancy Cartwright's philosophy of science*, Volume 3. Routledge.

Boyd, R. (1991). Realism, anti-foundationalism and the enthusiasm for natural kinds. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition 61*(1/2), 127–148.

Brandon, R. N. (1990). *Adaptation and environment.* Princeton University Press.

Brewer, W. F. and B. L. Lambert (2001). The theory-ladenness of observation and the theory-ladenness of the rest of the scientific process. *Philosophy of Science 68*(S3), S176–S186.

Bromberger, S. (1966). Why-Questions. In R. Colodny (Ed.), *Mind and Cosmos: Essays in Contemporary Science and Philosophy*, pp. 86–111. Pittsburgh, PA: University of Pittsburgh Press.

Brown, H. I. (1993). A theory-laden observation can test the theory. *The British Journal for the Philosophy of Science 44*(3), 555–559.

Brown, L. S. and G. Gabrielse (1986). Geonium theory: Physics of a single electron or ion in a Penning trap. *Reviews of Modern Physics 58*(1), 233.

Butterfield, J. (2014). Reduction, emergence, and renormalization. *The Journal of Philosophy 111*(1), 5–49.

Butterfield, J. and N. Bouatta (2015). Renormalization for Philosophers. In *Metaphysics in Contemporary Physics*, pp. 437–85. Leiden: Brill.

Callender, C. and J. Cohen (2006). There is no special problem about scientific representation. *Theoria 21*(1), 67–85.

Calmet, X. and F. Kuipers (2021). Theoretical bounds on dark matter masses. *Physics Letters B 814*, 136068.

Capdevila, B., A. Crivellin, S. Descotes-Genon, J. Matias, and J. Virto (2018). Patterns of New Physics in $b \to s\ell^+\ell^-$ transitions in the light of recent data. *Journal of High Energy Physics 2018*(1), 93.

Carnap, R. (1937). *Logical Syntax of Language.* London: Routledge.

Carnap, R. (1966). *Philosophical foundations of physics*, Volume 966. Basic Books New York.

Carnap, R. (1996). *Philosophy and Logical Syntax.* Bristol: Theoemmes.

Carnap, R. (2012). Empiricism, semantics, and ontology. In P. Benacerraf and H. Putnam (Eds.), *Philosophy of mathematics: Selected Readings* (2nd ed.)., pp. 241–257. Cambridge: Cambridge University Press.

Carnap, R. and A. George, R (1967). *The Logical Structure of the World.* Berkley, CA: University of California Press.

Carnap, R. and P. A. Schilpp (1963). *The Philosophy of Rudolf Carnap.* Cambridge University Press Cambridge.

Carney, D., S. Ghosh, G. Krnjaic, and J. M. Taylor (2020). Proposal for gravitational direct detection of dark matter. *Physical Review D 102*(7), 72003.

Carr, B., F. Kühnel, and M. Sandstad (2016). Primordial black holes as dark matter. *Physical Review D 94*(8), 83504.

Cartwright, N. (1983). *How the laws of physics lie.* Oxford: Clarendon Press.

Cartwright, N. (1999). *The dappled world: A study of the boundaries of science.* Cambridge University Press.

Cartwright, N., T. Shomar, and M. Suarez (1995). The tool box of science: Tools for the building of models with a superconductivity example. *Poznan Studies in the Philosophy of the Sciences and the Humanities 44*, 137–149.

Celis, A., J. Fuentes-Martin, M. Jung, and H. Serodio (2015). Family nonuniversal Z′ models with protected flavor-changing interactions. *Physical Review D 92*(1), 15007.

Chalmers, A. (1993). So the laws of physics needn't lie. *Australasian Journal of Philosophy 71*(2).

Clemencic, M., G. Corti, C. Easo, R. Jones, C, S. Milglioranzi, M. Pappagallo, and P. Robbe (2011). The LHCb Simulation Application, Gauss: Design, Evolution and Experience. *Journal of Physics: Conference Series 331*, 32023.

Clowe, D., M. Bradač, A. H. Gonzalez, M. Markevitch, S. W. Randall, C. Jones, and D. Zaritsky (2006). A direct empirical proof of the existence of dark matter. *The Astrophysical Journal Letters 648*(2), L109.

Contessa, G. (2010). Scientific models and fictional objects. *Synthese 172*(2), 215–229.

Craver, C. (2002). Structures of Scientific Theories. In K. Machamer, P and M. Silberstein (Eds.), *Blackwell Guide to the Philosophy of Science*, pp. 55–79. Oxford: Blackwell.

Craver, C. F. (2006). When mechanistic models explain. *Synthese 153*(3), 355–376.

Curtright, T. L., X. Jin, and C. K. Zachos (2012). Renormalization Group Flows, Cycles, and c-Theorem Folklore. *Physical review letters 108*(13), 131601.

da Costa, N. and S. French (2000). Models, Theories, and Structures: Thirty Years on. *Philosophy of Science 67*, S116–S127.

da Costa, N. and S. French (2003). *Science and partial truth: A unitary approach to models and scientific reasoning.* Oxford University Press.

De Baerdemaeker, S. (2021). Method-driven experiments and the search for dark matter. *Philosophy of Science 88*(1), 124–144.

de Lavallaz, A. and M. Fairbairn (2010). Neutron stars as dark matter probes. *Physical Review D 81*(12), 123521.

de Swart, J. (2020). Closing in on the Cosmos: Cosmology's Rebirth and the Rise of the Dark Matter Problem. In A. Blum, R. Lalli, and J. Renn (Eds.), *The Renaissance of General Relativity in Context*, pp. 257–284. Cham: Springer.

Dray, W. (1957). *Law and Explanation in History.* Oxford: Oxford University Press.

Duhem, P. (1954). *The Aim and Structure of Physical Theory.* Princeton, NJ: Princeton University Press.

Duncan, A. (2012). *The conceptual framework of Quantum Field Theory.* Oxford: Oxford University Press.

Dyson, F. J. (1952). Divergence of perturbation theory in quantum electrodynamics. *Physical Review 85*(4), 631.

Earman, J. (2004). Curie's principle and spontaneous symmetry breaking. *International Studies in the Philosophy of Science 18*(2-3), 173–198.

Earman, J. and D. Fraser (2006). Haag's theorem and its implications for the foundations of quantum field theory. *Erkenntnis 64*, 305–344.

Earman, J. and J. Roberts (1999). Ceteris paribus, there is no problem of provisos. *Synthese 118*(3), 439–478.

Edwards, P. N. (2010). *A vast machine: Computer models, climate data, and the politics of global warming.* MIT Press.

Einstein, A. (1905). On the electrodynamics of moving bodies. *Annalen der physik 17*(10), 891–921.

Einstein, A. (1988). *Ideas and Opinions.* New York: Crown Publishers.

Eklund, M. (2013). Carnap's metaontology. *Nous 47*(2), 229–249.

Elder, J. (2022). Black Hole Coalescence: Observation and Model.

Elliott, J, P. (1958). Collective Motions in the Nuclear Shell Model. *Proceedings of the Royal Society of London, Series A 245*, 128–45.

Elvira, V. D. (2017). Impact of detector simulation in particle physics collider experiments. *Physics Reports 695*, 1–54.

Eronen, M. I. (2015). Robustness and reality. *Synthese 192*(12), 3961–3977.

Evans, P. W. and K. P. Y. Thébault (2020). On the limits of experimental knowledge. *Philosophical Transactions of the Royal Society A 378*(2177).

Feng, J. L. (2010). Dark matter candidates from particle physics and methods of detection. *Annual Review of Astronomy and Astrophysics 48*, 495–545.

Ferenc, D., D. F. Šegedin, I. F. Šegedin, and M. Š. Ferenc (2019). Helium Migration through Photomultiplier Tubes–The Probable Cause of the DAMA Seasonal Variation Effect. *arXiv preprint arXiv:1901.02139*.

Finocchiaro, P. (2019). Ideology and its role in metaphysics. *Synthese*, 1–27.

Fletcher, S. C. (2020). The Principle of Stability. *Philosopher's Imprint 20*(3), 1–22.

Flocke, V. (2020). Carnap's Noncognitivism about Ontology. *Noûs 54*(3), 527–548.

Forster, M. R. (1988). Unification, explanation, and the composition of causes in Newtonian mechanics. *Studies In History and Philosophy of Science Part A 19*(1), 55–101.

Franklin, A. (1989). *The neglect of experiment.* Cambridge University Press.

Franklin, A. (1998). Selectivity and the Production of Experimental Results. *Archive for history of exact sciences 53*(5), 399–485.

Franklin, A. (2015). The Theory-Ladenness of Experiment. *Journal for General Philosophy of Science 46*(1), 155–166.

Franklin, A. (2018). On the renormalization group explanation of universality. *Philosophy of Science 85*(2), 225–248.

Franklin, A. (2020). Whence the Effectiveness of Effective Field Theories? *The British Journal for the Philosophy of Science 71*(4), 1235–1259.

Franklin, A., Anderson, Brock, Coleman, Downing, Gruvander, Lilly, Neal, Peterson, and Price (1989). Can a Theory-Laden Observation Test the Theory? *British Journal for the Philosophy of Science 40*(2), 229–231.

Franklin, A. and C. Howson (1984). Why do scientists prefer to vary their experiments? *Studies in History and Philosophy of Science Part A 15*(1), 51–62.

Fraser, D. (2009). Quantum field theory: Underdetermination, inconsistency, and idealization. *Philosophy of Science 76*(4), 536–567.

Fraser, J. D. (2020). The Real Problem with Perturbative Quantum Field Theory. *The British Journal for the Philosophy of Science 71*(2), 391–413.

Freeman, K. C. (1970). On the disks of spiral and S0 galaxies. *The Astrophysical Journal 160*, 811.

French, S. (2010). Keeping quiet on the ontology of models. *Synthese 172*(2), 231–249.

French, S. (2021). Identity conditions, idealisations and isomorphisms: a defence of the Semantic Approach. *Synthese 198*(24), 5897–5917.

Frigg, R. (2010). Models and fiction. *Synthese 172*(2), 251–268.

Frigg, R. and S. Hartmann (2016). Models in Science. In E. Zalta (Ed.), *Stanford Encyclopedia of Philosophy* (Spring 202 ed.).

Frigg, R. and J. Nguyen (2016). The fiction view of models reloaded. *Monist 99*(3), 225–242.

Frigg, R. and J. Nguyen (2017). Models and Representation. In L. Magnani and T. Bertolotti (Eds.), *Springer Handbook of Model-Based Science*. Cham: Springer.

Galison, P. (1987). *How experiments end*, Volume 88. University of Chicago Press Chicago.

Galison, P. (1997). *Image and logic: A material culture of microphysics.* University of Chicago Press.

Gaskins, J. M. (2016). A review of indirect searches for particle dark matter. *Contemporary Physics 57*(4), 496–525.

Gelfert, A. (2016). *How to do science with models : a philosophical primer.* Cham: Springer.

Gelfert, A. (2017). The Ontology of Models. In L. Magnani and T. Bertolotti (Eds.), *Springer Handbook of Model-Based Science.* Cham: Springer.

Giere, R. (1988). *Explaining Science: A cognitive approach.* Chicago: University of Chicago Press.

Giere, R. N. (2004). How models are used to represent reality. *Philosophy of science 71*(5), 742–752.

Gluscevic, V., Y. Ali-Haimoud, K. Bechtol, K. K. Boddy, C. Bœhm, J. Chluba, F.-Y. Cyr-Racine, C. Dvorkin, D. Grin, and J. Lesgourgues (2019). Cosmological probes of dark matter interactions: The next decade. *arXiv preprint arXiv:1903.05140*.

Glymour, B. (2000). Data and phenomena: A distinction reconsidered. *Erkenntnis 52*(1), 29–37.

Glymour, C. (2013). Theoretical equivalence and the semantic view of theories. *Philosophy of Science 80*(2), 286–297.

Godfrey-Smith, P. (2006). The strategy of model-based science. *Biology & Philosophy 21*(5), 725–740.

Goodman, J., M. Ibe, A. Rajaraman, W. Shepherd, T. M. P. Tait, and H.-B. Yu (2010). Constraints on dark matter from colliders. *Physical Review D 82*(11), 116010.

Greiner, W. and J. Maruhn (1996). *Nuclear models.* Berlin: Springer-Verlag.

Grüne-Yanoff, T. (2013). Appraising models nonrepresentationally. *Philosophy of Science 80*(5), 850–861.

Gueguen, M. (2020). On Robustness in Cosmological Simulations. *Philosophy of Science 87*(5), 1197–1208.

Hacking, I. (1983). *Representing and intervening: Introductory topics in the philosophy of natural science.* Cambridge university press.

Hacking, I. (2007). Natural kinds: Rosy dawn, scholastic twilight. *Royal Institute of Philosophy Supplements 61*, 203–239.

Hacking, I. (2009). *Scientific reason.* National Taiwan University Press Taipei.

Halvorson, H. (2012). What scientific theories could not be. *Philosophy of science 79*(2), 183–206.

Harkins, W. D. and E. D. Wilson (1915). Recent work on the structure of the atom.[third paper on atomic structure.]. *Journal of the American Chemical Society 37*(6), 1396–1421.

Harris, T. (2003). Data models and the acquisition and manipulation of data. *Philosophy of Science 70*(5), 1508–1517.

Hartmann, S. (1999). Models and Stories in Hadron Physics. In M. S. Morgan and M. Morrison (Eds.), *Models as mediators : perspectives on natural and social sciences*, pp. 326–346. Cambridge: Cambridge University Press.

Head, T. (2014). The LHCb trigger system. *Journal of Instrumentation 9*(09), C09015.

Hempel, C. (1958). The Theoritician's dilemma. In H. Feigl, M. Scriven, and G. Maxwell (Eds.), *Minessota Studies in the Philosophy of Science, vol.2*, pp. 37–98. Minneapolis: University of Minessota Press.

Hempel, C. G. (1965). *Aspects of scientific explanation*, Volume 1. Free Press New York.

Hempel, C. G. and P. Oppenheim (1948). Studies in the Logic of Explanation. *Philosophy of science 15*(2), 135–175.

Hey, S. P. (2015). Robust and discordant evidence: Methodological lessons from clinical research. *Philosophy of Science 82*(1), 55–75.

Huang, K. (2013). A critical history of renormalization. *International Journal of Modern Physics A 28*(29).

Hudson, R. (2009). The methodological strategy of robustness in the context of experimental wimp research. *Foundations of Physics 39*(2), 174–193.

Humair, T. (2019). *Testing lepton universality in penguin decays of beauty mesons using the LHCb detector.* Phd thesis, Imperial College London.

Humphreys, P. (2004). *Extending ourselves: Computational science, empiricism, and scientific method.* Oxford University Press.

Jacquart, M. (2021). Dark Matter and Dark Energy. In E. K. Knox and A. Wilson (Eds.), *Routledge Companion to the Philosophy of Physics*, pp. 731–743. New York: Routledge.

Jansson, L. and J. Tallant (2020). Quantitative parsimony: Probably for the better. *The British Journal for the Philosophy of Science*.

Kahlhoefer, F. (2017). Review of LHC dark matter searches. *International Journal of Modern Physics A 32*(13), 1730006.

Kaiser, D. (2009). *Drawing theories apart.* University of Chicago Press.

Karaca, K. (2013). The strong and weak senses of theory-ladenness of experimentation: Theory-driven versus exploratory experiments in the history of high-energy particle physics. *Science in Context 26*(1), 93–136.

Karaca, K. (2017). A case study in experimental exploration: exploratory data selection at the Large Hadron Collider. *Synthese 194*(2), 333–354.

Karaca, K. (2018). Lessons from the Large Hadron Collider for model-based experimentation: the concept of a model of data acquisition and the scope of the hierarchy of models. *Synthese 195*(12), 5431–5452.

Kinoshita, T. (2014). Tenth-order QED contribution to the electron g-2 and high precision test of quantum electrodynamics. *Proceedings of the conference in Honour of the 90th Birthday of Freeman Dyson*, 148–172.

Kitcher, P. (1989). Explanatory Unification and the Causal Structure of the World. In P. Kitcher and W. Salmon (Eds.), *Scientific Explanation*, pp. 410–505. Minneapolis,MN: University of Minessota Press.

Knuuttila, T. (2005). Models, representation, and mediation. *Philosophy of science 72*(5), 1260–1271.

Knuuttila, T. (2011). Modelling and representing: An artefactual approach to model-based representation. *Studies in History and Philosophy of Science Part A 42*(2), 262–271.

Knuuttila, T. and M. Boon (2011). How do models give us knowledge? The case of Carnot's ideal heat engine. *European journal for philosophy of science 1*(3), 309–334.

Kosso, P. (1989). *Observability and Observation in Physical Science.* Springer Netherlands.

Kosso, P. (2013). Evidence of dark matter, and the interpretive role of general relativity. *Studies in History and Philosophy of Science Part B - Studies in History and Philosophy of Modern Physics 44*(2), 143–147.

Kragh, H. (2012). *Niels Bohr and the quantum atom: The Bohr model of atomic structure 1913-1925.* OUP Oxford.

Kuhn, T. S. (1970). *The structure of scientific revolutions*, Volume 111. Chicago University of Chicago Press.

Kuorikoski, J., A. Lehtinen, and C. Marchionni (2010). Economic modelling as robustness analysis. *The British Journal for the Philosophy of Science 61*(3), 541–567.

Kusch, P. and H. M. Foley (1947). Precision Measurement of the Ratio of the Atomicg Values' in the P 3 2 2 and P 1 2 2 States of Gallium. *Physical Review 72*(12), 1256.

Kusch, P. and H. M. Foley (1948). The Magnetic Moment of the Electron. *Physical Review 74*(3), 250–263.

Lakatos, I. (1970). Falsification and the Methodology of Scientific Research Programmes. In I. Lakatos and A. Musgrave (Eds.), *Criticism and the Growth of Knowledge: Proceedings of the International Colloquium in the Philosophy of Science*, pp. 91–196. Cambridge: Cambridge University Press.

Lamb, W. E. and R. C. Retherford (1947). Fine Structure of the Hydrogen Atom by a Microwave Method. *Physical Review 72*(3), 241–243.

Landsman, N. P. (2013). Spontaneous symmetry breaking in quantum systems: Emergence or reduction? *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics 44*(4), 379–394.

Leclair, A., J. M. Román, and G. Sierra (2003). Russian doll renormalization group and Kosterlitz-Thouless flows. *Nuclear physics B 675*(3), 584–606.

Lehtinen, A. (2013). On the impossibility of amalgamating evidence. *Journal for general philosophy of science 44*(1), 101–110.

Lelli, F., S. S. McGaugh, and J. M. Schombert (2016). The small scatter of the baryonic Tully-Fisher relation. *The Astrophysical Journal Letters 816*(1), L14.

Leonelli, S. (2015). What counts as scientific data? A relational framework. *Philosophy of Science 82*(5), 810–821.

Leonelli, S. (2016). *Data-centric biology: A philosophical study.* University of Chicago Press.

Leonelli, S. (2019). What distinguishes data from models? *European journal for philosophy of science 9*(2), 22.

LHCb Collaboration (2008). The LHCb detector at the LHC. *Journal of instrumentation 3*(8), S08005.

LHCb Collaboration (2019). Search for Lepton-Universality Violation in $B^+ \to K^+\ell^+\ell^-$ Decays. *Physical Review Letters 122.*

LHCb Collaboration (2021). Test of lepton universality in beauty-quark decays. *arXiv preprint arXiv:2103.11769*.

Lionetto, F. (2018). *Measurements of Angular Observables of* $B^0 \to K^{*0}\mu^+\mu-$ *and* $B^0 \to K^{*0}e^+e-$ *Decays and the Upgrade of LHCb.* Phd thesis, University of Zurich.

Lisciandra, C. (2017). Robustness analysis and tractability in modeling. *European Journal for Philosophy of Science 7*(1), 79–95.

Lisovskyi, V. (2019). *Study of rare b-baryon decays and test of Lepton Universality at LHCb.* Phd thesis, Université Paris-Saclay.

Lista, L. (2016). *Statistical methods for data analysis in particle physics.* Springer.

Lloyd, E. A. (1994). *The structure and confirmation of evolutionary theory.* Princeton University Press.

Lloyd, E. A. (2010). Confirmation and robustness of climate models. *Philosophy of Science 77*(5), 971–984.

Lloyd, E. A. (2015). Model robustness as a confirmatory virtue: The case of climate science. *Studies in History and Philosophy of Science Part A 49*, 58–68.

Luković, V., P. Cabella, and N. Vittorio (2014). Dark matter in cosmology. *International Journal of Modern Physics A 29*(19), 1443001.

Lutz, S. (2017). What was the syntax-semantics debate in the philosophy of science about? *Philosophy and Phenomenological Research 95*(2), 319–352.

Lyons, L. (Ed.) (2006). *Statistical Problems in Particle Physics, Astrophysics and Cosmology.* Oxford: Imperial College Press.

Macarthur, D. and H. Price (2007). Pragmatism , Quasi-realism and the Global Challenge. *New Pragmatists 91*, 91–121.

Mach, E. (2013). *The Science of Mechanics: A Critical and Historical Exposition of its Principles.* Cambridge: Cambridge University Press.

Mach, E. (2014). On the Economical Nature of Physical Inquiry. In E. Mach and T. J. McCormack (Eds.), *Popular Scientific Lectures*, Cambridge Library Collection - Physical Sciences, pp. 186–213. Cambridge: Cambridge University Press.

Machamer, P., L. Darden, and C. F. Craver (2000). Thinking about mechanisms. *Philosophy of science 67*(1), 1–25.

Margenau, H. (1935). Methodology of modern physics. *Philosophy of Science 2*(2), 164–187.

Mari, L. and A. Giordani (2014). Modelling Measurement: Error and Uncertanty. In M. Boumans, G. Hon, and A. Petersen (Eds.), *Error and Uncertainty in Scientific Practice*, pp. 79–96. London: Pickering & Gatto.

Martens, N. C. M. (2022). Dark Matter Realism. *Foundations of Physics 52*(1), 1–19.

Martens, N. C. M. and D. Lehmkuhl (2020a). Cartography of the space of theories: an interpretational chart for fields that are both (dark) matter and spacetime. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics 72*, 217–236.

Martens, N. C. M. and D. Lehmkuhl (2020b). Dark Matter = Modified Gravity? Scrutinising the spacetime − matter distinction through the modified gravity/dark matter lens. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics 72*, 237–250.

Massey, R., T. Kitching, and J. Richard (2010). The dark matter of gravitational lensing. *Reports on Progress in Physics 73*(8), 86901.

Massimi, M. (2007). Saving unobservable phenomena. *British Journal for the Philosophy of Science 58*(2), 235–262.

Massimi, M. (2018). Perspectival Modeling. *Philosophy of Science 85*(3), 335–359.

Mathieu, V., N. Kochelev, and V. Vento (2009). The physics of glueballs. *International Journal of Modern Physics E 18*(1), 1–49.

Mauri, A. (2019). *Direct and Indirect Searches for New Physics in* $b \to s\ell\ell$ *Decays*. Phd thesis, University of Zurich.

Mayo, D. G. (1996). *Error and the growth of experimental knowledge.* University of Chicago Press.

McAllister, J. W. (1997). Phenomena and patterns in data sets. *Erkenntnis 47*(2), 217–228.

McCoy, C. D. and M. Massimi (2018). Simplified models: a different perspective on models as mediators. *European Journal for Philosophy of Science 8*(1), 99–123.

McGaugh, S. (2020). Predictions and outcomes for the dynamics of rotating galaxies. *Galaxies 8*(2), 35.

McMullin, E. (1978). Structural explanation. *American Philosophical Quarterly 15*(2), 139–147.

McMullin, E. (1985). Galilean idealization. *Studies in History and Philosophy of Science Part A 16*(3), 247–273.

Meinong, A. (1960). The Theory of Objects. In M. Chisolm, R (Ed.), *Realism and the Background of Phenomenology.* Glencoe, IL: Free Press.

Meinong, A. (1968). Über Moglichkeit und Wahrscheinlichkeit [On Possibility and Probability]. In R. Haller and R. Kindinger (Eds.), *Alexius Meinong Gesamtausgabe.*

Menn, W., O. Adriani, G. C. Barbarino, G. A. Bazilevskaya, R. Bellotti, M. Boezio, E. A. Bogomolov, L. Bonechi, M. Bongi, and V. Bonvicini (2013). The PAMELA space experiment. *Advances in space research 51*(2), 209–218.

Meynell, L. (2008). Why Feynman diagrams represent. *International Studies in the Philosophy of Science 22*(1), 39–59.

Miller, M. E. (2018). Haag's theorem, apparent inconsistency, and the empirical adequacy of quantum field theory. *British Journal for the Philosophy of Science*.

Miller, M. E. (2021a). Infrared Cancellation and Measurement. *Philosophy of Science 88*(5), 1125–1136.

Miller, M. E. (2021b). Mathematical structure and empirical content. *British Journal for the Philosophy of Science*.

Morgan, M. S. and M. Morrison (1999). *Models as mediators : perspectives on natural and social sciences.* Cambridge: Cambridge University Press.

Morgante, E. (2018). Simplified dark matter models. *Advances in High Energy Physics 2018.*

Mormann, T. (2007). The Structure of Scientific Theories in Logical Empiricism. In A. Richardson and T. Uebel (Eds.), *The Cambridge Companion to Logical Empiricism*, pp. 136–162. Cambridge: Cambridge University Press.

Morozov, A. and A. J. Niemi (2003). Can renormalization group flow end in a big mess? *Nuclear Physics B 666*(3), 311–336.

Morris, S. (2018). Carnap and quine: Analyticity, naturalism, and the elimination of metaphysics. *Monist 101*(4), 394–416.

Morrison, M. (1999). Models as autonomous agents. In M. Morgan and M. Morrison (Eds.), *Models as mediators : perspectives on natural and social sciences.* Cambridge: Cambridge University Press.

Morrison, M. (2007). Where have all the theories gone? *Philosophy of Science 74*(2), 195–228.

Morrison, M. (2015). *Reconstructing reality : models, mathematics, and simulations.* Oxford studies in philosophy of science. New York: Oxford University Press.

Morrison, M. and M. S. Morgan (1999). Models as mediating instruments. In M. Morrison and M. S. Morgan (Eds.), *Models as Mediators*, pp. 10–37. Cambridge: Cambridge University Press.

Muller, K. (2019). Tests of Lepton Flavour Universality at LHCb. *Journal of Physics, Conference Series 1271*.

Natarajan, A. (2013). Bounds on Dark Matter from CMB Observations. In *Sources and Detection of Dark Matter and Dark Energy in the Universe*, pp. 67–71. Springer.

Needham, P. (1991). Duhem and Cartwright on the Truth of Laws. *Synthese 89*(1), 89–109.

Nelson, R. A. and M. G. Olsson (1986). The pendulum - Rich physics from a simple system. *American Journal of Physics 54*(2), 112–121.

Nguyen, J. (2020). It's Not a Game: Accurate Representation with Toy Models. *The British Journal for the Philosophy of Science TA - TT - 71*(3), 1013–1041.

Nolan, D. (1997). Quantitative parsimony. *British Journal for the Philosophy of Science 48*(3).

Norton, J. D. (2012). Approximation and idealization: Why the difference matters. *Philosophy of Science 79*(2), 207–232.

Odenbaugh, J. and A. Alexandrova (2011). Buyer beware: robustness analyses in economics and biology. *Biology and Philosophy 26*(5), 757–771.

Oort, J. H. (1932). The force exerted by the stellar system in the direction perpendicular to the galactic plane and some related problems. *Bulletin of the Astronomical Institutes of the Netherlands 6*, 249.

Parker, W. S. (2011). When climate models agree: The significance of robust model predictions. *Philosophy of Science 78*(4), 579–600.

Parker, W. S. (2017). Computer simulation, measurement, and data assimilation. *The British Journal for the Philosophy of Science 68*(1), 273–304.

Peirce, C. (1940). *The philosophy of Peirce: Selected writings.* London: Routledge and Kegan Paul.

Perszyk, K. J. (1993). *Nonexistent objects: Meinong and contemporary philosophy*, Volume 49. Springer Science & Business Media.

Peskin, M. and D. Schroeder (1995). *An Introduction To Quantum Field Theory.* Reading, MA: Addison-Wesley.

Pincock, C. (2007). Mathematical Idealization. *Philosophy of Science 74*(5), 957–967.

Popper, K. (1963). *Conjectures and refutations: The growth of scientific knowledge.* routledge.

Portides, D. (2005). Scientific Models and the Semantic View of Scientific Theories. *Philosophy of Science 72*(5), 1287–1298.

Portides, D. (2021). Idealization and abstraction in scientific modeling. *Synthese 198*(24), 5873–5895.

Price, H. (2004). Naturalism Without Representationalism. In M. De Caro and D. Macarthur (Eds.), *Naturalism in Question*, pp. 71–88. Cambridge, MA: Harvard University Press.

Price, H. (2007). Quining Naturalism. *Journal of Philosophy 104*(8), 375–402.

Price, H. (2009). Metaphysics After Carnap : The Ghost Who Walks? *Metametaphysics: New Essays on the Foundations of Ontology*, 320–346.

Psillos, S. (2011). Living with the abstract: realism and models. *Synthese 180*(1), 3–17.

Quine, W. V. (1951a). Main Trends in Recent Philosophy: Two Dogmas of Empiricism. *The Philosophical Review 60*(1), 20–43.

Quine, W. V. (1951b). On Carnap's views on ontology. *Philosophical Studies 2*(5), 65–72.

Quine, W. V. (1956). Quantifiers and propositional attitudes. *The Journal of Philosophy 53*(5), 177–187.

Quine, W. V. (1960). *Word and object.* Cambridge, MA: MIT Press.

Quine, W. V. (1992). Structure and Nature. *The Journal of Philosophy 89*(1), 5–9.

Rappoccio, S. (2019). The experimental status of direct searches for exotic physics beyond the standard model at the Large Hadron Collider. *Reviews in Physics 4*.

Redhead, M. (1980). Models in physics. *British Journal for the Philosophy of Science 31*(2), 145–163.

Reichert, M. and J. Smirnov (2020). Dark matter meets quantum gravity. *Physical Review D 101*(6), 63015.

Reutlinger, A. (2014). Why is there universal macrobehavior? Renormalization group explanation as noncausal explanation. *Philosophy of Science 81*(5), 1157–1170.

Reutlinger, A., D. Hangleiter, and S. Hartmann (2020). Understanding (with) Toy Models. *The British Journal for the Philosophy of Science*.

Ritson, S. and K. Staley (2020). How uncertainty can save measurement from circularity and holism. *Studies in History and Philosophy of Science Part A*.

Rivat, S. (2019). Renormalization scrutinized. *Studies in History and Philosophy of Science Part B - Studies in History and Philosophy of Modern Physics 68*, 23–39.

Ruben, D.-H. (2015). *Explaining explanation.* Routledge.

Rubin, V. and K. Ford (1970). Rotation of the Andromeda nebula from a spectroscopic survey of emission regions. *The Astrophysical Journal 159*, 379.

Ruetsche, L. (2011). *Interpreting quantum theories.* Oxford University Press.

Saatsi, J. and A. Reutlinger (2018). Taking reductionism to the limit: How to rebut the antireductionist argument from infinite limits. *Philosophy of Science 85*(3), 455–482.

Salmon, W. C. (1971). *Statistical explanation and statistical relevance*, Volume 69. University of Pittsburgh Press.

Salmon, W. C. (1984). *Scientific explanation and the causal structure of the world.* Princeton University Press.

Salmon, W. C. (1989). *Four Decades of Scientific Explanation.* Minneapolis: University of Minessota Press.

Sanders, R. H. and S. S. McGaugh (2002). Modified Newtonian dynamics as an alternative to dark matter. *Annual Review of Astronomy and Astrophysics 40*(1), 263–317.

Schaffner, K. F. (1969). The Watson-Crick model and reductionism. *The British Journal for the Philosophy of Science 20*(4), 325–348.

Schelling, T. C. (1971). Dynamic models of segregation. *Journal of mathematical sociology 1*(2), 143–186.

Schindler, S. (2013). Theory-laden experimentation. *Studies in History and Philosophy of Science Part A 44*(1), 89–101.

Schroeder, M. (2010). *Noncognitivism in Ethics.* New York: Routledge.

Schumann, M. (2019). Direct detection of WIMP dark matter: concepts and status. *Journal of Physics G: Nuclear and Particle Physics 46*(10), 103003.

Schupbach, J. N. (2018). Robustness analysis as explanatory reasoning. *The British Journal for the Philosophy of Science 69*(1), 275–300.

Schwartz, M. D. (2014). *Quantum field theory and the standard model.* Cambridge University Press.

Schwinger, J. (1948). On quantum-electrodynamics and the magnetic moment of the electron. *Physical Review 73*(4), 416–417.

Shapere, D. (1982). The concept of observation in science and philosophy. *Philosophy of science 49*(4), 485–525.

Sider, T. (2009). Ontological realism. In D. Chalmers, M. David, and R. Wasserman (Eds.), *Metametaphysics*, pp. 384–423. Oxford: Oxford University Press.

Sinervo, P. K. (2003). Definition and treatment of systematic uncertainties in high energy physics and astrophysics. *Statistical Problems in Particle Physics, Astrophysics, and Cosmology*, 122–129.

Sklar, L. (2000). *Theory and truth : philosophical critique within foundational science.* Oxford: Oxford University Press.

Smeenk, C. (2020). Some reflections on the structure of cosmological knowledge. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics 71*, 220–231.

Smeenk, C. and S. C. Gallagher (2020). Validating the Universe in a Box. *Philosophy of Science 87*(5), 1221–1233.

Staley, K. W. (2020). Securing the empirical value of measurement results. *The British Journal for the Philosophy of Science 71*(1), 87–113.

Stegenga, J. (2009). Robustness, discordance, and relevance. *Philosophy of Science 76*(5), 650–661.

Stein, H. (1989). Yes, but... Some Skeptical Remarks on Realism and Anti-Realism. *Dialectica 43*(1-2), 47–65.

Steinle, F. (1997). Entering new fields: Exploratory uses of experimentation. *Philosophy of science 64*, S65–S74.

Stöltzner, M. (2017). Feynman diagrams as models. *The Mathematical Intelligencer 39*(2), 46–54.

Suárez, M. (2004). An inferential conception of scientific representation. *Philosophy of science 71*(5), 767–779.

Suárez, M. and N. Cartwright (2008). Theories: Tools versus models. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics 39*(1), 62–81.

Suppe, F. (1977). *The Structure of Scientific Theories.* Urbana, IL: University of Illinois Press.

Suppe, F. (1989). *The semantic conception of theories and scientific realism.* University of Illinois Press.

Suppe, F. (2000). Understanding scientific theories: An assessment of developments, 1969-1998. *Philosophy of science 67*, S102–S115.

Suppes, P. (1957). *Introduction to Logic.* Princeton: D. Van Nostrand Co.

Suppes, P. (1960). A comparison of the meaning and uses of models in mathematics and the empirical sciences. *Synthese 12*(2-3), 287–301.

Suppes, P. (1962). Models of Data. In E. Nagel, P. Suppes, and A. Tarski (Eds.), *Logic, methodology and philosophy of science: Proceedings of the 1960 international congress*, pp. 252 – 261. Stanford: Stanford University Press.

Suppes, P. (1967). What is a scientitic theory? *Philosophy of science today*, 55–67.

Suppes, P. (2002). *Representation and Invariance of Scientific Structures.* Stanford, CA: CSLI Publications.

Tal, E. (2017a). A Model-Based Epistemology of Measurement. In N. Mößner and A. Nordman (Eds.), *Reasoning in measurement*, pp. 233–253. Taylor & Francis.

Tal, E. (2017b). Measurement in Science. In N. Zalta, Edward (Ed.), *Stanford Encyclopedia of Philosophy* (Fall 2017 ed.).

Tarski, A. (1953). A General Method in Proofs of Undecidability. In *Studies in Logic and the Foundations of Mathematics*, Volume 13, pp. 1–34. Elsevier.

Tauber, J. A. (2004). The Planck mission. *Advances in Space Research 34*(3), 491–496.

Teller, P. (2001). Twilight Of The Perfect Model Model. *Erkenntnis 55*(3), 393–415.

Teller, P. (2018). Making worlds with symbols. *Synthese*.

Teubert, F. (2016). Rare Decays Probing Physics Beyond the Standard Theory. In L. Maiani and L. Rolandi (Eds.), *The Standard Theory of Particle Physics: Essays to Celebrate CERN's 60th Anniversary.*, pp. 357–370. World Scientific.

Thomasson, A. L. (2006). Debates about the Ontology of Art: What are We Doing Here? *Philosophy Compass 1*(3), 245–255.

Thomasson, A. L. (2014). *Ontology made easy.* Oxford University Press.

Thomson-Jones, M. (2010). Missing systems and the face value practice. *Synthese 172*(2), 283–299.

Thomson-Jones, M. (2017). Against Bracketing and Complacency. In M. Slater and Z. Yudell (Eds.), *Metaphysics and the Philosophy of Science*, pp. 229–251. New York: Oxford University Press.

Toon, A. (2010). The ontology of theoretical modelling: Models as make-believe. *Synthese 172*(2), 301–315.

Toon, A. (2012). *Models as Make-Believe: Imagination, Fiction and Scientific Representation.* Palgrave Macmillan UK.

Van Dokkum, P., S. Danieli, Y. Cohen, A. Merritt, A. J. Romanowsky, R. Abraham, J. Brodie, C. Conroy, D. Lokhorst, and L. Mowla (2018). A galaxy lacking dark matter. *Nature 555*(7698), 629–632.

van Fraassen, B. C. (1970). On the extension of Beth's semantics of physical theories. *Philosophy of science 37*(3), 325–339.

van Fraassen, B. C. (1980). *The scientific image.* Oxford: Clarendon Press.

van Fraassen, B. C. (1989). *Laws and Symmetry.* Oxford University Press.

van Fraassen, B. C. (2008). *Scientific Representation: Paradoxes of Perspective.* Oxford: Clarendon Press.

van Fraassen, B. C. (2014). One or two gentle remarks about Hans Halvorson's critique of the semantic view. *Philosophy of Science 81*(2), 276–283.

Van Inwagen, P. (2020). The neo-Carnapians. *Synthese 197*(1), 7–32.

van Roojen, M. (2018). Moral Cognitivism Vs. Non-Cognitivism. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2018 ed.).

Vanderburgh, W. L. (2003). The Dark Matter Double Bind: Astrophysical Aspects of the Evidential Warrant for General Relativity. *Philosophy of Science 70*(4), 812–832.

Vanderburgh, W. L. (2005). The methodological value of coincidences: Further remarks on dark matter and the astrophysical warrant for general relativity. *Philosophy of Science 72*(5), 1324–1335.

Vanderburgh, W. L. (2014). On the interpretive role of theories of gravity and 'ugly' solutions to the total evidence for dark matter. *Studies in History and Philosophy of Science Part B - Studies in History and Philosophy of Modern Physics 47*(1), 62–67.

Verhaegh, S. (2017). Blurring Boundaries: Carnap, Quine, and the Internal-External Distinction. *Erkenntnis 82*(4), 873–890.

Verhaegh, S. (2018). *Working from Within: The Nature and Development of Quine's Naturalism.* Oxford: Oxford University Press.

Vickers, P. (2013). *Understanding Inconsistent Science.* Oxford: Oxford University Press.

Virdee, T. S. (2016). Beyond the standard model of particle physics. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 374*(2075).

Wallace, D. (2006). In defence of naiveté: The conceptual status of Lagrangian quantum field theory. *Synthese 151*, 33–80.

Wallace, D. (2018). The quantum theory of fields. In E. Knox and A. Wilson (Eds.), *The routledge companion to the philosophy of physics*. Routledge.

Wanke, R. (2016). How to deal with systematic uncertainties. *Data Analysis in High Energy Physics*, 263–296.

Wartofsky, M, W. (1979). Models, Metaphysics and the Vagaries of Empiricism. In *Models. Boston Studies in the Philosophy of Science, vol 48*. Dordrecht: Springer.

Weisberg, M. (2006). Robustness analysis. *Philosophy of science 73*(5), 730–742.

Weisberg, M. (2007). Three kinds of idealization. *The journal of Philosophy 104*(12), 639–659.

Weisberg, M. (2013). *Simulation and similarity.* New York: Oxford University Press.

Weisberg, M., M. Jacquart, B. Madore, and M. Seidel (2018). The Dark Galaxy Hypothesis. *Philosophy of Science 85*(5), 1204–1215.

Williams, P. (2015). Naturalness, the autonomy of scales, and the 125GeV Higgs. *Studies in History and Philosophy of Science Part B - Studies in History and Philosophy of Modern Physics 51*, 82–96.

Williams, P. (2018). Renormalization group methods. In E. Knox and A. Wilson (Eds.), *The routledge companion to the philosophy of physics*. Routledge.

Wilson, K. G. (1971). Renormalization group and strong interactions. *Physical Review D 3*(8), 1818.

Winther, R. G. (2012). Interweaving categories: Styles, paradigms, and models. *Studies in History and Philosophy of Science Part A 43*(4), 628–639.

Winther, R, G. (2021). The Structure of Scientific Theories. *The Stanford Encyclopedia of Philosophy*.

Woodward, J. (1989). Data and phenomena. *Synthese 79*(3), 393–472.

Woodward, J. (2005). *Making things happen: A theory of causal explanation.* Oxford university press.

Woodward, J. (2006). Some varieties of robustness. *Journal of Economic Methodology 13*(2), 219–240.

Worrall, J. (2011). Underdetermination, realism and empirical equivalence. *Synthese 180*(2), 157–172.

Worrall, J. (2014). Prediction and accommodation revisited. *Studies in History and Philosophy of Science Part A 45*, 54–61.

Wüthrich, A. (2010). *The genesis of Feynman diagrams*, Volume 26. Springer Science & Business Media.

Yablo, S. (1998). Does Ontology rest on a mistake? *Aristotelian Society Supplementary Volume 72*(1), 229–261.

Zahar, E. (1973). Why did Einstein's programme supersede Lorentz's?(I). *The British Journal for the Philosophy of Science 24*(2), 95–123.

Zee, A. (2010). *Quantum field theory in a nutshell* (2nd ed.). Princeton, NJ: Princeton University Press.

Zwicky, F. (1933). Die rotverschiebung von extragalaktischen nebeln. *Helvetica physica acta 6*, 110–127.