Cosmology: The Impossible Integration?

by

Siska De Baerdemaeker

B.A., Philosophy, KU Leuven, 2012

M.A. Philosophy, KU Leuven, 2013

B.Sc. Physics, KU Leuven, 2014

Submitted to the Graduate Faculty of

the Dietrich School of Arts and Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2020

UNIVERSITY OF PITTSBURGH

DIETRICH SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

Siska De Baerdemaeker

It was defended on

March 31st 2020

and approved by

John Norton, History and Philosophy of Science, University of Pittsburgh

Arthur Kosowsky, Physics and Astronomy, University of Pittsburgh

Sandra Mitchell, History and Philosophy of Science, University of Pittsburgh

James Woodward, History and Philosophy of Science, University of Pittsburgh

Dissertation Director: John Norton, History and Philosophy of Science, University of Pittsburgh

Cosmology: The Impossible Integration?

Siska De Baerdemaeker, PhD

University of Pittsburgh, 2020

My dissertation introduces a new account of how empirical methods and lines of evidence can come to bear on cosmological model-building. Through a careful study of the recent history of cosmology and dark matter research, I explicate a new type of justification for experiments, a 'method-driven logic'. This structure of justification underlies terrestrial experiments researching dark matter and dark energy, but it is more generally prevalent in cases of an underdescribed target. Using a method-driven logic comes with a cost, however. Specifically, interpreting the empirical results of experiments justified through a method-driven logic is non-trivial: negative results warrant secure constraints on the space of possibilities for the target, whereas significant positive results remain ambivalent. While this ambivalence can be resolved through the amalgamation of multiple lines of evidence, this solution is sometimes faced with conflicts between those lines of evidence. I propose that, under specific circumstances, restricting the relevant empirical evidence can be warranted. Finally, I discuss the use of cosmological evidence as a constraint in other subfields of physics. This brings me full-circle on the integration of disciplines in cosmology—an integration driven by experimental practice.

Keywords: History and philosophy of science, integration, philosophy of experiment, cosmology, dark matter.

Table of Contents

List of Figures

Preface

After finishing this work of six years, I am indebted to a large group of people who helped me achieve this. First, my thanks goes out to my advisor, John Norton, for taking me under his wing. John, your feedback, guidance, and encouragements have been instrumental these past six years. It has been an incredible learning experience to work with you.

I also thank my committee. They are all giants in the field, and I feel privileged to have studied with them. Thank you to Sandy Mitchell, for helping me sharpen my ideas, to Jim Woodward, for all of his precise feedback, and to Arthur Kosowsky, for being open to take on a philosopher and for being a patient teacher about cosmology.

I am also grateful to mentors outside of the HPS community, including Allan Franklin, who first sparked my interest in experimental practice, and Michela Massimi, whose philosophical work and mentorship are a continuous source of inspiration. A further word of thanks goes to everyone in the broader philosophy of cosmology community, for stimulating discussions and feedback at various events in the past six years.

The unsung heroes of the HPS department are the administrative staff, past and present. They also deserve a word of thanks. Natalie Schweninger once literally saved my personal identity. Joann McIntyre always had a word of encouragement or a sugary pick-me-up ready. Katie Labuda was a steadfast support through the final steps of finishing a degree in the midst of unprecedented circumstances. Matt Ceraso and Diana Volkar deserve a word of thanks, as well as the staff from the Center for Philosophy of Science who helped organize Super-PAC: Joyce MacDonald, Carolyn Oblak, and Cheryl Greer.

Throughout my time in Pittsburgh, I was lucky enough to join the HPS graduate community. I would like to thank the HPS graduate students, past and present, for being a welcoming support network, an intellectual inspiration, and the best group of friends (and for giving me occasion to wear a blow-up crown and a sash!). Thank you to Aaron Novick, Brendan Fleig-Goldstein, David Colaço, Dasha Pruss, Dejan Makovec, Evan Pence, Gal Ben Porath, Jen Whyte, Joe McCaffrey, JP Gamboa, Lauren Ross, Katie Creel, Katie Morrow, Marina Baldissera Pacchetti, Marina DiMarco, Michael Begun, Michal De-Medonsa, Mike

Miller, Morgan Thompson, Nedah Nemati, Nuhu Osman Attah, Shahin Kaveh, and Willy Penn. A special thank you goes out to Haixin Dang, Mahi Hardaloupas, Jacob Neal and Zina Ward. Finally, I am grateful to call Nora Boyd my mentor, conference co-organizer, collaborator, and all-around fairy godmother.

I also want to thank everyone outside of academia who supported me. Thank you to my roommates of the past five years, India Foy and Caroline Lafferty. Thank you to Kevin Baud for the past year. And thank you to my Belgian friends, who stuck with me through this adventure: Alixe Tsang-Leclerq, Andries Verslyppe, Béatrice Ranwez, Bram Vaassen, Célestine Maes, Eléonore Moncheur, Jan Doumont, Lore Dours, Marieke Brugnera, Mattias Verbeeck, Ruth Heirman, Sander Glas, Sarah Jacobs, Sixtine van Outryve, and Wai-Hoi Tsang.

The final word of thanks goes to my family. Thank you to my parents, Josse De Baerdemaeker and Katelijne Van Overwalle, and to my sisters, Marieke, Trees and Klara. I have been lucky to follow in all of your footsteps. I dedicate this dissertation to them, and to the memory of Fons Meskens, a world-class philosopher in his own right.

I defended this dissertation in the midst of the Covid-19 pandemic. I especially dedicate this piece to everyone working to keep others safe during this global health crisis, from healthcare workers to grocery store clerks. The importance of their work eclipses anything in the current dissertation.

1.0 Introduction

Cosmology is currently in a peculiar situation. On the one hand, it has developed a successful phenomenological model to reconstruct the evolution of the universe at the largest scales to a very high degree of accuracy. That so-called concordance model of cosmology, ΛCDM,[1] describes the formation of the lightest elements, the decoupling of the cosmic microwave background radiation (CMB) and the formation of large-scale structure through gravitational collapse, all in an expanding universe. ΛCDM sits at the intersection of general relativity, particle physics and astrophysics, and it posits genuinely novel physics in the form of dark energy and dark matter. Cosmology generally integrates models and lines of evidence of those disciplines in its pursuit to reconstruct the evolution of the universe—with success. When the Planck Collaboration (2018, p. 59) completed the most recent analysis of their observations of the CMB, they concluded that they "find that the base-ΛCDM model provides a remarkably good fit to the Planck power spectra and lensing measurements". The success of ΛCDM was recognized most recently with the 2020 Nobel Prize in Physics being awarded to Jim Peebles, for his contributions to the development of the concordance model.

On the other hand, the fundamental nature of approximately 95% of the current total energy density of the universe remains unknown. That is, based on observations of the CMB and the large-scale structure of the universe, amongst others, cosmologists have come to the conclusion that the universe today is dominated by dark energy and dark matter, approximately 68.3% and 26.8% of the total current energy density,[2] respectively. What dark matter and dark energy consist of, remains to be discovered. For dark matter, proposals range from various Weakly Interacting Massive Particles (WIMPs), over axions to sterile

[1]In this abbreviation, Λ is the cosmological constant, and CDM refers to Cold Dark Matter. See Weinberg 2008 for an introduction to cosmology. See Ellis 2018 for a summary of ΛCDM and various issues with the concordance model.

[2]These percentages reflect the current (i.e. redshift $z = 0$) fractional contributions to the total energy density ρ_0 of the universe. They are usually represented in terms of the dimensionless energy density paramater $\Omega_0 = \frac{\rho_0}{\rho_c}$, where ρ_c is the critical density for which the geometry of spacetime is (extremely close to) flat (represented by the curvature parameter $k = 0$). Observations indicate that $\rho_0 = \rho_c$, such that $\Omega_0 = 1 = \Omega_{\Lambda,0} + \Omega_{m,0} + \Omega_{k,0}$, where $\Omega_{\Lambda,0} = 0.6889 \pm 0.0056$, $\Omega_{m,0} = 0.3147 \pm 0.0074$, and $\Omega_{k,0} = 0$ (values from Planck Collaboration 2018). These percentages have changed over the expansion history of the universe.

neutrinos (see Bertone, Hooper, and Silk 2005; Particle Data Group 2019b for an overview). For dark energy, the space of possibilities is even broader; proposals include vacuum energy density fluctuations, chameleon fields, or a so-called quintessence (see Particle Data Group 2019a for an overview).

In an attempt to find the fundamental nature of dark matter (and to a lesser extent dark energy), the last four decades have seen an ever-increasing involvement of other disciplines[3] to generate novel lines of evidence about both (see De Swart, Bertone, and Van Dongen 2017 for a historical account of dark matter research). This is an instantiation of a general tendency in cosmology to integrate models, experimental methods and lines of evidence from other disciplines. In and of itself, the fact that cosmology is such an integrative discipline generates several epistemological challenges that parallel existing philosophical work in philosophy of the life sciences (cf. also infra). These challenges are exacerbated by the fact that, in cosmology, one is faced with the genuinely novel physics of dark matter and dark energy. Given that little is known about dark matter and dark energy, outside of their role in ΛCDM, it is not immediately obvious how those other disciplines can contribute to dark matter and dark energy-research—yet they do.

In this dissertation, I raise three epistemological challenges about empirical evidence in cosmology that arise at the intersection of cosmology and particle physics:

1. Dark matter and dark energy are supposedly genuinely novel, that is, they are not described by any existing theory of fundamental physics. Current experiments use techniques that have proven to be effective in probing known fundamental physics to probe this novel physics. How can the effectiveness of these experiments be established? Do they need to make specific assumptions about the genuinely novel physics, and if so, what are these assumptions?
2. Given that various lines of evidence in cosmology result from experiments that make substantive assumptions about their target, do these assumptions undermine the conclusions about the target that these lines of evidence warrant? And can these results be

[3] Here, 'discipline' is not used in any substantive way related to questions about demarcation. I solely use it as a collective name for a set of research strategies, experiments and theoretical models. For example, particle physics is a discipline where the standard model of particle physics is central, which is usually explored through experiments in particle accelerators like the Large Hadron Collider.

compared with one another, if they make different assumptions about the target?

3. With accelerator experiments reaching their limits, particle physics has started to exploit the early universe as a new source of evidence. This is a substantive methodological shift. What epistemic shift is associated with this methodological shift, if any?

I take up these challenges in the next four chapters.

Chapter 2.0 takes up the first challenge of generating lines of evidence in context of contemporary dark matter research. Here, I focus specifically on the justification for particle dark matter searches. I argue that various experiments in contemporary dark matter research have been justified through a so-called 'method-driven logic'. The two characterizing features of a method-driven logic are, first, that an existing empirical method is repurposed to a new target. Second, the theoretical model of that target is adapted to incorporate the features that the target would need to have in order for the existing empirical method to be effective in probing that target. In other words, the theoretical model of the target is adapted to a particular method, rather than, as is more common, that a method is chosen based on its fit with a particular target. I show that this method-driven logic is particularly useful (and used!) whenever the existing description of a target is limited. As an example, I show how this method-driven logic is used in the justification of various terrestrial particle searches for dark matter. Thus, the method-driven logic helps to explain certain unexpected connections between disciplines, connections that are ubiquitous in modern cosmology.

Illuminating the method-driven logic brings a new challenge to the forefront about how the resulting data should be interpreted. In chapter 3.0, I show that a significant positive result from a method-driven experiment turns out to be ambivalent as to what conclusions about the target it warrants, while a negative result warrants secure constraints on the space of possibilities for the target. Not all is lost to ambivalence when a method-driven experiment leads to a positive results: I show how various strategies (paralleling Staley (2004)'s security of evidence framework) allow for a possible escape out of the ambivalence. Once again, this solution also presents a new challenge: making results from method-driven experiments comparable with one another, despite the different assumptions about the target they might make.

I broaden up my discussion of comparing lines of evidence in chapter 4.0 by focusing on conflicts between lines of evidence. Stegenga (2009) refers to this as the 'problem of inconsistent evidence' in context of the robustness debate in philosophy of science. The challenge of inconsistent evidence extends beyond the robustness debate: it poses a challenge to the basic requirement of empirical adequacy of scientific models and theories with regards to all the available empirical evidence. In context of cosmology, such inconsistencies sometimes arise between lines of evidence with their causal origin in the evolution of the universe, and lines of evidence that originate from some other target than that evolution at the largest scale. I show that, under specific circumstances, and in case of conflict between the two types of evidence, it can be reasonable to (temporarily) ignore the second type of empirical evidence.

Finally, in chapter 5.0, I turn towards the third challenge. Since cosmology sits at the cross-roads of various disciplines, it provides new potential for these disciplines to further their own research programmes. This is particularly important for particle physics, where physics beyond the standard model remains elusive and terrestrial particle accelerators are nearing their capacity limits. Cosmological probes constitute a methodological shift for particle physics from accelerator experiments to reconstructing observations of the early universe. I investigate the locus—if there is any—of the associated epistemic shift. I argue that that epistemic shift does not lie in the shift from experiment to observation, nor in the shift from direct observation to inference based on effects. Rather, it is in a propagation of errors from the lines of evidence informing cosmological theorizing, all the way back to the cosmological constraints on physics beyond the standard model.

Throughout the next four chapters, several central themes will emerge. A first theme is that of integration. My focus on the intersection of various disciplines in context of cosmology is reminiscent of the notion of integration that has been extensively discussed in philosophy of the life sciences. There, it was initially developed as an alternative view on inter-theory relations to reduction or unification by Darden and Maull (1977). Their motivation was to provide a philosophical analysis of the development of models of complex target systems. They argued that modeling complex systems requires the integration of various theoretical frameworks, and that these inter-theory relations could not be adequately described by the existing accounts of inter-theory relations.

Since that first proposal, various approaches to successful integration have been identified (see Mitchell 2003; Love 2008; Love and Lugar 2013; O'Malley and Soyer 2012; Leonelli 2016; Currie 2018, among many others). The broad view that underlies all these approaches is that science becomes integrative whenever one discipline is insufficient to make progress on a (complex or multi-level) research problem. This implies, as Potochnik (2011) argues, that a large portion of scientific research is integrative. Cosmology is no different, due to the scope and scale of cosmology's central research problem ('how did the universe evolve from an initial hot dense state to its current state?'). Given the widespread prevalence of integrative practices in scientific research, this dissertation does not aim to provide a unified account of 'integration'. Nonetheless, questions about integration will return—specifically with a focus on the integration of various empirical methods and the resulting lines of evidence (also the topic of Leonelli 2016; Mitchell and Gronenborn 2017; Currie 2018, for example). The arguments from the next four chapters will contain some general lessons for the role of integration in science. I will return to them in the concluding chapter of this dissertation.

Second, since I discuss the generation of and integration of multiple lines of evidence, I will also touch on the notion of robustness, and more specifically on robust measurement results, or triangulation. Following Wimsatt (2007)'s general account, an empirical results is considered robust if that result is, first, invariant for a range of different measurement procedures, and where, second, any failure of invariance can be explained. If an empirical result is found to be robust, it is often inferred that the result is tracking an actual physical effect, rather than some unaccounted for source of error. In this dissertation, I raise a new challenge for robustness arguments in context of dark matter searches (chapter 2.0), but I also use some of the existing literature on strengthening robustness arguments as a guideline for strengthening the results of those dark matter searches (chapter 3.0).

Finally, by its focus on empirical evidence in observational cosmology, my dissertation contributes to a revival and reorientation in philosophy of cosmology towards observational cosmology. This revival has in part been inspired by recent research projects led by Michela Massimi ("Perspectival Realism. Science, Knowledge, and Truth from a Human Vantage Point") and by Chris Smeenk and Jim Weatherall ("New Directions in Philosophy of Cosmology"). Before this reorientation, philosophers of cosmology focused primarily on theoretical issues, such as global underdetermination and the cosmological principle (Earman

1995; Beisbart and Jung 2006; Beisbart 2009; Butterfield 2014), anthropic reasoning (Barrow 1993; McMullin 1993; Cirkovic 2002; Weinstein 2006; Benétreau-Dupin 2015), typicality (McCoy 2015; McCoy 2017), or the status of (eternal) inflation (Earman and Mosterintl 1999; Tegmark 2003; Kragh 2009; Kragh 2013; Brandenberger 2014; Smeenk 2014). In the past five years, observational cosmology has started to receive similar scrutiny. Some of this more recent work has focused on the empirical evidence supporting dark matter and Modified Newtonian Dynamics, or MOND (Massimi 2019; Merrit 2017), the history of dark matter (De Swart, Bertone, and Van Dongen 2017), computer simulations in cosmology (Gueguen 2020; Smeenk and Gallagher 2020) and the nature of observations in cosmology and astrophysics (Weisberg et al. 2018). I can only hope that this trend will continue, and that the present work contributes positively towards that goal.

2.0 Method-Driven Experiments and the Search for Dark Matter[1]

The cosmological concordance model posits the presence of dark matter based on its gravitational effects on galaxy-, cluster-, and cosmological scales. The same observations that provide evidence for the existence of dark matter also show that dark matter is not constituted by anything studied by high-energy physics so far. This makes it puzzling that so many dark matter particle searches using common experimental techniques from high-energy physics are under way. This chapter illuminates the structure of justification underlying the dark matter searches.

2.1 INTRODUCTION

It is widely accepted in philosophy of science that having multiple lines of evidence, and multiple independent lines of evidence for a hypothesis, model or parameter value is preferable, and in some cases necessary. Cosmologists have taken this adage to heart. In the last three decades, they have relied on various types of astrophysical and cosmological observations, as well as terrestrial experiments to produce evidence constraining the cosmological concordance model, ΛCDM.

The relevance of these terrestrial experiments to cosmology is not always obvious. For example, there are experiments underway that use experimental methods from high-energy particle physics to investigate dark matter (see below), or technology from atomic physics to investigate dark energy (see for example Hamilton et al. 2015). Meanwhile, all available astrophysical and cosmological evidence supporting dark matter indicates that, if constituted by a (class of) particle(s), it is unlike any other particle studied by high-energy particle physics so far. Dark energy fares worse: although there is extensive evidence from cosmological observations for its effect on the expansion of the universe, its other properties are a

[1]A slightly modified version of this chapter has been accepted for publication in *Philosophy of Science* on 11/08/2019; all modifications were merely for consistency with the rest of the dissertation.

complete mystery. Current theorizing about dark matter and dark energy, aside from their effects on cosmological and astrophysical scales, excels in the negative: many possibilities have been excluded, none have been supported by empirical evidence.

Given that it is widely accepted that dark matter and dark energy are fundamentally different from any particle or entity that high-energy physics or atomic physics have studied in the past, it is puzzling that methods from these respective disciplines are employed to learn more about their properties. How can these experiments be justified? How do particle physicists argue that their experimental methods will be effective in probing dark matter, despite dark matter not being constituted by any particles in the current standard model of particle physics?

In this chapter, my primary goal is to answer this question of justification: I expose a new logic to structure the justification for a particular method choice. This logic differs from the logic that is often assumed, in that it is *method-driven*, rather than target-driven. I argue that the method-driven logic plays a crucial role when knowledge of the target is minimal, as is the case for dark matter or dark energy. Exposing the method-driven logic brings to the forefront some questions about the availability of robustness arguments and methodological pluralism more generally. A secondary goal of the chapter is to begin unearthing these questions in context of dark matter searches.

I begin by providing necessary terminological clarification as ground-work (section 2.2). I then explain the common target-driven logic of method choice and I contrast it with the less familiar method-driven logic in section 2.3. Section 2.4 contains a discussion of contemporary dark matter research as a detailed illustration of the latter. Before concluding, I draw implications for methodological pluralism in section 2.5.

2.2 METHODS, TARGET SYSTEMS, AND THEIR FEATURES

In order to get a better handle on this method-driven logic, it will be useful to take a step back and determine what constitutes a particular method. By ‘method’, I do not mean anything akin to a ‘unified scientific method’ like Mill’s methods or the logical positivists’ hypothetico-deductivism.[2] 'Method' here refers to something much more specific: a method is any activity that generates empirical evidence, where that activity can be applied in various research contexts and to various target systems. A method should be describable such that it is transferable across target systems, and specific enough such that any misapplication can be identified.

The use of the term 'activity' in the definition is purposefully vague: methods span a wide range of scientific practices, across disciplines and scales. Often, they will appear as (sets of) protocols in the 'method'-sections of scientific research papers. Perhaps the most useful characterization here comes in the form of some paradigmatic examples: a method can be anything from radiometric dating for fossils, to using particle colliders and assorted data processing to search for new elementary particles, or from using optical telescopes to observe sunspots and the moons of Jupiter to randomized controlled clinical trials in medicine. All these activities or clusters of activities consist of a protocol that uses some empirical input to produce data and ultimately a line of evidence for a particular phenomenon.[3] 'Conducting an experiment on a target system' does not qualify as a specific method, however, since it is too general to specify in what contexts this method can reliably to a target, and in what contexts it can not.

Regardless of the specifics of the protocol, a method should be applicable in various research contexts and to various target systems.[4] A method should be such that it is not tied down to one specific phenomenon or research context, even if the method in practice appears to be tied to a unique event. For example, while cosmologists can only observe one cosmic microwave background (CMB), their methods—using specific types of telescopes or radio receivers, among others—used in the mapping of the CMB are transferable to other targets.

The goal of the types of methods under consideration here, finally, is to generate empirical data that can be used as evidence for hypotheses about the target system. Whether or not

[2]See Andersen and Hepburn 2016 for an overview of attempts at defining a unified scientific method. For discussions about how the notion of 'integration', which originates from the life sciences, differs from and potentially replaces this unity of science-view, see Wylie 1999; O'Malley and Soyer 2012.

[3]I deliberately leave out methods of theoretical scientists. Although my views may extend to them, I believe they are different enough from empirical methods to not include them in the current chapter.

[4]See Norton 2018 for further discussion of the transferability of methods, specifically in context of Einstein's development of General Relativity.

that is good evidence will depend on how the method was applied to the target system and how the subsequent data processing happened—'using a scientific method' does not imply success.

A second term in need of clarification is that of a 'target system'. The target system is that system in the world about which a scientist's research aims to generate knowledge. The scientist's goal is to explore and model the features and causal interactions of the target system.[5]

At first sight, this definition may strike one as circular: wouldn't a scientist only know what they are investigating after they have finished the investigation? For example, didn't researchers at CERN only know what the 125 GeV-Higgs boson was after they had discovered it? The circularity is only apparent. It is true that some definition of the target system needs to be accepted before an investigation of the target system can commence, just like particle physics already had a hypothetical description of the Higgs boson and its role in the standard model before the discovery. That description or local theory can be minimal and independent of the new features that a particular experiment is investigating. Nonetheless, it still plays a crucial role in the justification of method choice, as I will show below.

To clarify the cluster of concepts, consider the following examples. A current high-profile set of experiments in physics attempts to measure the neutron lifetime (see Yue et al. 2013; Pattie et al. 2018). The target system is the neutron as described by nuclear physics. This description includes an estimate of the neutron mass, size, and magnetic moment, as well as the feature of nuclear β-decay, the process that determines the mean lifetime of the neutron τ_n (the precise value of τ_n affects the helium-to-hydrogen ratio generated by Big Bang Nucleosynthesis in the early universe, and it might help constrain extensions to the standard model of particle physics—hence its importance). Two types of methods are used to measure τ_n: one traps neutrons in a gravito-magnetic 'bottle' for an extended period of time, and counts the decay products after various time intervals. Another method uses a

[5]This term is in part adopted from the literature on external validity of experimentation (see for example Guala 2003) and the literature on simulations (see for example Parker 2009; Winsberg 2009; Parke 2014). The target system is contrasted with the laboratory system, experimental system or object system, i.e. the actual system on which experiments are being conducted. These two categories do not always come apart, as is the case for at least some of the experiments I discuss below. However, even in cases where they do come apart, I take the arguments below to still apply. I am grateful to an anonymous referee for pushing me on this point.

focused neutron beam and traps and counts decay products along the beam line.

Mitchell and Gronenborn (2017) provide an example from biology. Current molecular biology aims to determine the folding structure of various proteins. In this case, the target systems are the various proteins, and their feature of interest is their folding structure. Mitchell and Gronenborn describe how that folding structure can be measured through nuclear magnetic resonance or X-ray crystallography, and in what ways the two methods are independent from one another—despite sharing some common assumptions about the protein's primary structure, the sequence of amino acids.

2.3 CHOOSING THE RIGHT METHOD

How is a particular method choice justified? In other words, given a particular target, how do scientists argue that a (set of) method(s) will be effective when trying to learn more about the target and its features? Or, when multiple methods are available and scientists can only run a limited number of experiments, how do they argue that one (set of) method(s) will be more effective when trying to learn more about the target and its features than another (set)?

I propose that two types of logic of method choice can be at play in answering these questions: a target-driven logic, and a method-driven logic. (The use of 'logic' here merely indicates that there is a general schema that can be codified and applied broadly.) Both logics can be used to justify why a particular method can be used to construct a new line of evidence for a particular feature of a target system. The former, more common logic relies primarily on pre-existing knowledge of the target to justify the method choice. The latter, less familiar logic is prevalent in situations where this pre-existing knowledge is sparse and can therefore not be relied upon in the justification. For the purpose of this chapter, I will assume that the methods under consideration are well-developed and that their common sources of error are generally known, as this is the case for the dark matter experiments discussed in section 2.4.

2.3.1 Target-driven method choice

Very often, method choice follows a target-driven logic, where the justification relies on previous knowledge of the target as the prime differentiating factor between methods. "Knowledge" here indicates that the target and its various features are described by an established, empirically confirmed scientific theory. Target-driven method choice adheres to the following structure:[6]

> Given a target system T, with known features A and B.
> Method 1 uses feature A of possible targets to uncover potential new feature X.
> Method 2 uses feature B of possible targets to uncover potential new feature X.
> Method 3 uses feature C of possible targets to uncover potential new feature X.

> The preferred methods to uncover a potential new feature X of T will be methods 1 and 2, while method 3 remains out of consideration (for now).

In short, when a given method adheres to a feature that the target is known to have, it will be preferred over methods that adhere to features the target either does not have, or for which it is unclear whether the target has them. Of course, since the accepted theory about the target can evolve, the specifics that are substituted for 'method 1', or 'feature A' can also evolve over time—hence the bracketed caveat in the conclusion.

To see this abstract logic in action, consider again the two examples from the previous section. First, the neutron lifetime experiments:

> Given the neutron, which decays through nuclear β-decay, which has mass and a magnetic moment, and which has a radius of approximately $0.8 \times 10^{-15}m$.
> Bottle-experiments trap ultra-cold neutrons through their mass and magnetic moment and count the remaining neutrons that have not yet undergone β-decay at various times t to derive τ_n from the exponential decay function $N(t) = N_{t=0}e^{-t/\tau_n}$.
> Beam-experiments use a focused beam of cold neutrons that moves continuously through a proton trap, and count the number of neutrons N in a well-defined volume of the beam, as well as the rate of neutron decay dN/dt (based on the β-decay products (protons) that get periodically ejected from that volume) to derive τ_n from the differential decay function $dN/dt = -N/\tau_n$.
> Optical microscopy uses a minimal resolution of the order of $10^{-7}m$ to magnify optical features of small objects.

[6]For the purpose of this chapter, I will focus on possible experiments that all try to probe the same feature X of the target. I believe my arguments extend straightforwardly to cases where different methods probe different features as well.

Bottle and beam experiments can both be used to measure τ_n. Optical microscopy uses length scales much larger than the size of a neutron, and can therefore not be used to determine an individual neutron's properties.

Although the example of the optical microscope is contrived, my hope is that it conveys the structure of target-driven logic here: methods are selected based on pre-existing knowledge of the target, and on whether the methods in question employ features the target is known to have.

Along with neutron lifetime experiments, the target-driven logic of method choice also applies to the various protein folding structure experiments:

Given proteins, which are made up of atoms that have a nucleus and an electron cloud.
NMR experiments use the magnetic moment of atomic nuclei to determine the nuclei's, and therefore the atoms' positions.
X-ray crystallography uses the diffraction of X-rays by electron clouds to determine the electron clouds' density, and therefore the atoms' positions.

NMR and X-ray crystallography can both be used to determine the folding structure of proteins.

Again, particular methods are chosen based on pre-existing knowledge of the target system. It is the pre-existing knowledge of the target that determines whether or not an established method can reasonably be expected to be effective for investigating a new feature X or not.

2.3.2 Method-driven method choice

The above target-driven method choice may be a familiar one, but it is not the only logic of method-choice. Scientists are sometimes confronted with target systems where the definition is so thin that the target-driven logic cannot be employed to justify particular experimental explorations of new features of the target system. For example, in case of dark energy, its only commonly accepted features are that it is a majority contribution to the current energy-density of the universe and that it causes the universe's expansion to accelerate. In the words of the Dark Energy Task Force:

> Although there is currently conclusive observational evidence for the existence of dark energy, we know very little about its basic properties. It is not at present possible, even with the latest results from ground and space observations, to determine whether a cosmological constant, a dynamical fluid, or a modification of general relativity is the correct explanation. We cannot yet even say whether dark energy evolves with time. (Albrecht et al. 2006, p. 1)

Without a well-developed, empirically confirmed theory of dark energy that describes its basic properties, it is impossible to apply the target-driven logic explained in the previous section. There are not enough known features to latch on to to uncover new features of the target system.

While the target-driven logic cannot be employed, it is not the case that scientists are left completely in the dark and cannot justify their choice of methods. Rather, they employ a different logic: a *method-driven logic*. This method-driven logic has a similar set-up as the target driven-logic, namely:

> Given a target system T.
> Method 1 uses feature A of possible targets to uncover new feature X.
> Method 2 uses feature B of possible targets to uncover new feature X.
> Method 3 uses feature C of possible targets to uncover new feature X.

However, unlike for the target-driven logic, it is now not stated from the beginning that the target T has features A, B or C that are used by methods 1, 2 or 3 to uncover a new feature X. Thus, the next step in the method-driven logic needs to be different:

> If T has feature A, method 1 can be used to uncover new feature X.
> If T has feature B, method 2 can be used to uncover new feature X.
> If T has feature C, method 3 can be used to uncover new feature X.

With this different set-up compared to the method-driven logic, another premise is needed to complete the justification of the method-choice: which of the antecedent clauses in the above set of premises can plausibly be triggered?

> It is possible and plausible that T has features A and B, but it is either impossible or implausible (but possible) that T has feature C.

> The preferred methods to uncover a potential new feature X will be methods 1 and 2, while method 3 remains out of consideration (for now).

In the method-driven logic, the justification of the method-choice primarily appeals to what features the target *would need to have* in order for various established methods to be effective in probing its features. Rather than appealing to an established theory of the target, the justification primarily appeals to pre-existing knowledge of the available methods. This guides what assumptions about the target need to be made for that method to produce reliable evidence about various features of the target.

Using the method-driven logic in a responsible manner requires that these assumptions are plausible (and at minimum possible), but it does not require sufficient empirical evidence to accept the assumption as 'known' or empirically well-confirmed. It is at this point that the pre-existing knowledge of the target system T still plays a role in context of the method-driven logic. In the most ideal situation, the pre-existing knowledge allows scientists to construct a plausibility argument for T having a specific feature, turning the assumptions from an 'allowed' to an 'educated' guess, at best. But at minimum, the pre-existing knowledge of T delineates what additional assumptions about the target are possible, and which ones are already excluded. For example, dark energy cannot be any known form of matter, since it counteracts gravity. Similarly, dark matter cannot be constituted entirely by standard model particles. Any method that uses those respective features is therefore already excluded for dark energy or dark matter research.

One important qualification to the discussion is in order here.[7] So far, I have introduced the distinction between the target- and the method-driven logic as a dichotomy. However, the two logics lie on a continuous spectrum. It is possible to reformulate the general structure of the target-driven logic such that it is almost identical with the method-driven logic, except for the final premise. Rather than it being possible or plausible that the target has a particular assumed feature, this premise would read that there is very high confidence that the target has a particular assumed feature (which I indicated as 'known' in section 2.3.1). The cases I discuss here all lie on one or another extreme of the spectrum—this allows me to showcase important implications of the method-driven logic for the interpretation of the results (see section 2.5).

Finally, let me also briefly touch on the relation between my distinction here, and the

[7] I would like to thank an anonymous referee for this point.

literature on exploratory experimentation. Exploratory experiments are commonly defined as contrasting with confirmatory experiments: while confirmatory experiments aim to test a particular hypothesis, exploratory experiments do not (see for example Franklin 2005). It is sometimes argued that they are particularly attractive for discovering new phenomena whenever theoretical frameworks are in turmoil or underdeveloped—something I have also indicated as a reason to use a method-driven logic.

I take my distinction between target- and method-driven logic to be orthogonal to the confirmatory/exploratory experimentation distinction, however. I follow Karaca (2017) and Colaço (2018) in that theory of the target system and the method plays a role in confirmatory and exploratory experiments alike. Colaço (2018, p. 38) explicitly identifies "determining that the system is an appropriate candidate for the application of a technique" as one role for theory in context of exploratory experiments. My discussion here focuses specifically on how this task is fulfilled in light of that available theory, whereas the difference between confirmatory and exploratory experiments focuses on the further *aim* of the resulting experiment (confirmatory experiments aim to *test* the theory of the target, whereas exploratory experiments do not). It is therefore entirely possible for exploratory experiments to be target-driven (I take Colaço's cases from brain research to be examples of this) or method-driven (the dark matter example discussed below might qualify as such).

Before discussing the implications of the use of the method-driven logic, more needs to be said to elucidate this method-driven logic itself. What does it mean to 'make assumptions plausible' without having empirical evidence for them? What can scientists conclude, if anything at all, if their experiments using methods 1, or 2, or both lead to a positive result? And what if the contrary is the case? To answer these and other questions about the method-driven logic further, I now turn towards contemporary dark matter research.

2.4 METHOD-DRIVEN CHOICES IN DARK MATTER RESEARCH

The method-driven logic underlies several research projects in cosmology, including different searches for dark matter particles. I first explain why dark matter was introduced, and how the reasons for its introduction define dark matter as a target. This definition is thin, however, and provides very little in terms of properties of dark matter to use as in the target-driven logic above. Instead, particle physicists have justified dark matter searches with a method-driven logic.

2.4.1 Defining the target system of dark matter

An abundance of cosmological and astrophysical evidence supports the presence of an additional, non-baryonic matter component contributing to the energy density of the universe. The first observations in support of dark matter date back to the 1930s, to Zwicky's observations of velocity dispersions in the Coma Cluster (Zwicky 2009). In the 1980s, Rubin's observations of flat galaxy rotation curves became the first widely accepted source of evidence for dark matter (Rubin and Ford Jr. 1970; Rubin, Ford Jr., and Rubin 1973). Other evidence comes from lensing events like the Bullet Cluster (Clowe et al. 2006), weak lensing surveys, theories of cosmological structure formation (White and Rees 1978) and the Cosmic Microwave Background (CMB) anisotropy power spectrum (Planck Collaboration 2018).

In the course of the past four decades, it has become abundantly clear that astrophysical and cosmological evidence only gets scientists so far in learning more about dark matter. The available evidence supports dark matter's gravitational effects, and puts the dark matter contribution to the current energy density of the universe at approximately 26%. It also constrains what the particle properties of the constituents of dark matter can be: dark matter is non-baryonic (i.e. it is not constituted by particles in the standard model of particle physics), its coupling to standard model particles through the strong or electromagnetic interaction is very limited or non-existing, and its self-interaction cross-section is limited. It is this small set of properties that defines dark matter as a target: a form of non-baryonic matter that acts gravitationally and where there are strong upper limits on various possible couplings to standard model particles, as well as on its self-interaction cross-section.

That definition is remarkably thin. For all that the cosmological and astrophysical evidence reveals about what features dark matter can *not* have, it does not indicate much about what the particle properties of dark matter are. It does not reveal what the coupling mechanism of dark matter particles to standard model physics are. Worse, it does not even reveal whether there is such a coupling in the first place.

Nonetheless, the thin definition plays a crucial role in dark matter research for three reasons. First, the different dark matter experiments described below all share this common definition of the target. This makes it at least possible that, despite their different search strategies and further assumptions about dark matter, all experiments are probing the same target (although I will raise some issues with this in section 2.5). Relatedly, the thin definition constrains the space of possibilities for any further theorizing about dark matter: any more elaborate model for dark matter particles better adhere to the properties above, if it wants to claim to describe dark matter. Models proposing standard model neutrinos cannot describe dark matter, for example, since cosmological structure formation excludes neutrinos as a credible dark matter candidate. Finally, the definition of dark matter provides some resources for formulating possibility and plausibility arguments that can help guide the method-driven research.

Within the constraints from cosmology and astrophysics, a variety of experiments has been proposed and executed in the search for dark matter. These range from searches at accelerators like the Large Hadron Collider (LHC), over looking for nuclear recoils of dark matter particles with heavy atomic nuclei, to finding signatures of annihilation products in astrophysical observations. The candidate particle favoured by most so far is the Weakly Interacting Massive Particle (WIMP), a (class of) particle(s) with a mass of the order of O(100 GeV) coupling through the weak interaction.[8] This because of the so-called "WIMP-miracle": including WIMPs in a standard Big Bang Nucleosynthesis-scenario would automatically obtain the dark matter abundance derived from observations of the CMB and structure formation. WIMPs are therefore considered plausible dark matter candidates (although failure to turn up any positive detection result has recently started to put pressure on the WIMP hypothesis). Regardless of what happens to the WIMP hypothesis in the future, it is useful to examine in detail how physicists have tried to find WIMPs.

For the present purposes, I will focus on two sets of experiments: current production

[8]This definition in itself is still quite minimal; many particle candidates would fit within this definition, including several supersymmetric particles.

searches at the LHC, and the early days of direct detection experiments (see also figure 2.1). Both approaches have had to make additional assumptions about the particle properties of dark matter without having empirical evidence for them. The most basic assumption is depicted in figure 2.1: dark matter particles (χ) must couple to standard model particles (f) through some mediator (dashed line). How that coupling is modelled and exploited varies between the different experimental approaches.

2.4.2 Production experiments

Production experiments, mostly conducted at the LHC, aim to produce dark matter through collisions of standard model particles in an accelerator.[9] The hope is that dark matter will be another successful LHC search to uncover previously unobserved physics, similar to the discovery of the Higgs boson, and the general methodology behind it is very similar. The LHC looks for elementary particles by colliding proton beams at very high energy levels. These collisions generate approximately 600 million collisions per second, of which only a small fraction is recorded for further analysis.[10] After going through trigger systems, data reductions and background reductions, new physics presents itself as a detected excess signal, or as excess missing energy, above the background. A discovery is claimed if the excess can be calculated as having a high enough statistical significance, and if all known sources of systematic error have been excluded to a satisfactory degree.

Two types of dark matter searches are conducted at the LHC. The first type searches for particles described by specific extensions of the standard model, for example supersymmetry (SUSY). The hope is that there is a supersymmetric particle that also constitutes dark matter. If this assumption is correct, then SUSY particles, and therefore dark matter, can be found at the LHC (specifically by searching for a final state signature that is rich in jets and that has a significant amount of missing transverse energy).

The second type of searches is more general and less dependant on proposed extensions of the standard model of particle physics. Rather than focusing on constraining the properties

[9] The discussion below is based largely on Hong 2017 and Buchmueller, Doglioni, and Wang 2017. I refer the reader there for more details.

[10] See Karaca 2017 for a detailed discussion of event selection procedures and data processing at the LHC, in context of exploratory searches for physics beyond the standard model.

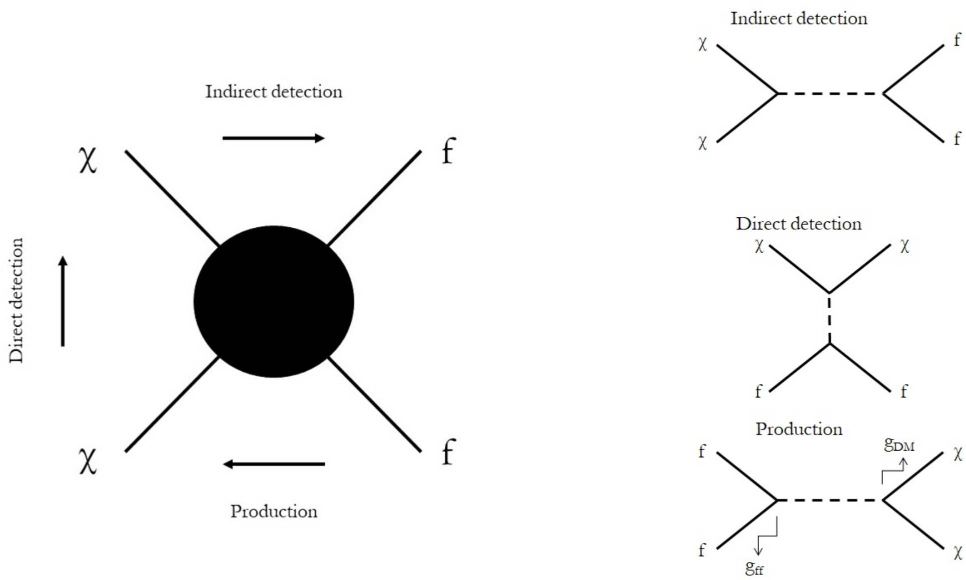

Figure 2.1: A schematic overview of three types of searches for dark matter particles (represented by χ), all based on interactions with standard model particles (represented by f). The three Feynman diagrams on the right are specifications of the diagram on the left, where the arrows on the left diagram indicate the direction in which the diagram should be read. On the right, the dotted lines represent the mediator, and g_{ff} and g_{DM} are the coupling strenghts at the standard model and dark matter vertex, respectively.

of the dark matter particle itself, these searches focus on the mediator for the coupling of the ‘dark sector’ (the collection of all dark matter particles to standard model particles). One possible mediator is the Higgs boson: current constraints on the branching ratios of the different Higgs decay channels still allow for a significant ‘invisible Higgs’-sector, leaving room for possible couplings to the dark sector.

These so-called ‘model-independent’ searches are not tied to the actual production of dark matter—they could just as well detect the mediator particle based on its decay into the standard model particles.[11] The decay products of the mediator can either be (1) missing energy, if it decays into dark matter, or (2) unaccounted for but detected standard model particles. The goal of the model-independent searches is primarily to discover an unknown mediator to physics beyond the standard model, and to exploit this new gateway to subsequently search for dark matter candidates. One benefit of this approach is that it potentially mitigates the worry about distinguishing dark matter from neutrinos, another particle species presenting itself as missing energy.

Production experiments thus instantiate the method-driven logic as follows.

> Given dark matter as described and defined in section 2.4.1.
> Accelerator experiments use the coupling of new particle species to standard model particles to detect new particles or couplings in the form of excess events above the background.
> If dark matter is made up of SUSY particles, or it couples to standard model particles through a mediator (invisible Higgs or otherwise), accelerator experiments can be used to search for signatures of dark matter particles or dark matter mediators.
> It is possible and plausible that dark matter is made up of SUSY particles, or that it couples to standard model particles through a mediator (invisible Higgs or otherwise).

> Accelerator experiments can be used to search for signatures of dark matter particles or dark matter mediators.

Important here is that there is no independent evidence for SUSY particles constituting dark matter, or even for the coupling between dark matter and standard model particles. Again, based on the cosmological and astrophysical evidence for dark matter, it is entirely possible that dark matter does not couple to standard model particles at all. Accelerator experiments would be useless in the search for dark matter in that case: they require some

[11] Model-independent searches at the LHC are therefore also not necessarily tied to the WIMP-hypothesis. Exploring the details of this is beyond the scope of the current chapter, however.

non-gravitational coupling to standard model particles to detect a signature. Luckily, the cosmological and astrophysical evidence does not exclude non-gravitational coupling either. Independent arguments in favor of SUSY- or invisible Higgs-physics then provide a plausibility argument in favor of the assumptions that are required to make accelerator experiments effective dark matter probes.

2.4.3 Direct detection experiments

My brief discussion of production experiments provided a first example of how the method-driven logic can be implemented in practice. They are not the sole type of search currently under way; direct detection experiments are another. The early days of direct detection experiments also provide a nice example of scientific practice appealing to the method-driven logic.

Direct detection searches like LUX, CoGeNT, CRESST, or CDMS look for signals of dark matter particles that scatter off heavy nuclei like xenon. The basic principle behind the experiments is the same as that behind neutrino searches: a scattering event results in recoil energy being deposited, which can be transformed into a detectable signal. Similarly, direct detection searches look for evidence of a scattering event of a dark matter particle off a heavy nucleus. According to the WIMP hypothesis, that scattering happens through the weak interaction. Should a WIMP scatter off one of the xenon nuclei in the detector, it would deposit some recoil energy which can be detected using scintillators and photomultiplier tubes, for example.

Even from this brief summary, it becomes clear that the justification for direct detection experiments is following a method-driven logic. The justification runs as follows:

> Given dark matter as described and defined in section 2.4.1.
> Neutrino-detection type experiments use the weak coupling of (extra-terrestrial) particles to detect their presence by their deposited energy in scattering events of those particles from atomic nuclei.
> If dark matter particles are weakly interacting, have a mass of the order of 100 GeV, and exist stably in the galactic halo, detectors similar to those used for neutrino detection can be used to search for signatures of dark matter particles existing in the halo of the Milky Way.
> It is plausible that dark matter particles are weakly interacting, have a mass of the order of 100 GeV, and exist stably in the galactic halo.

Detectors similar to those used for neutrino detection can be used to search for signatures of dark matter particles existing in the halo of the Milky Way.

To fully explore the justification for the various premises and the argument structure as a whole, let me return to the early days of direct detection experiments, in the late 1980s. Of particular interest here is a review article by Primack, Seckel, and Sadoulet (1988) that summarizes arguments for the effectiveness of direct detection experiments.

First, the authors use astrophysical and cosmological evidence to establish that dark matter exists, and that it must be non-baryonic (the first premise above). The authors appeal to flat galaxy rotation curves as evidence for dark matter.[12] The non-baryonic nature of dark matter is then established by first appealing to galaxy formation theories and Big Bang Nucleosynthesis constraints on the baryonic energy density component of the universe, and then comparing this limit to (admittedly very weak) constraints on the total energy density of the universe.

The second premise, on neutrino detection techniques, is not explicitly addressed in the review paper. Instead, they refer to previous papers on neutrino detectors by Goodman and Witten (1985) and Wasserman (1986). The two papers both first describe neutrino-detectors and subsequently argue that those neutrino-detectors might be effective for dark matter as well. For example, Goodman and Witten write:

> Dark galactic halos may be clouds of elementary particles so weakly interacting or so few and massive that they are not conspicuous. [...] Recently, Drukier and Stodolsky proposed a new way of detecting solar and reactor neutrinos. The idea is to exploit elastic neutral-current scattering of nuclei by neutrinos. [...] The principle of such a detector has already been demonstrated. In this paper, we will calculate the sensitivity of the detector [...] to various dark-matter candidates. (Goodman and Witten 1985, p. 3059)

Similarly, Wasserman writes:

> Recently, a new type of neutrino detector, which relies on the idea that even small neutrino energy losses ($\Delta E_\nu \gtrsim 1keV$) in cold material ($T \backsim 1 - 10mK$) with a small specific heat

[12] The authors caution in their review that Modified Newtonian Dynamics (MOND) could provide an alternative explanation for the flat rotation curves. This is to be expected given the history of dark matter: most of the current evidence that favours the existence of dark matter over MOND, like the Bullet Cluster or weak lensing surveys, was discovered in later decades.

> could produce measurable temperature changes, has been proposed. [...] The purpose of this paper is to examine the possibility that such a detector can be used to observe heavy neutral fermions ($m \gtrsim 1GeV$) in the Galaxy. Such particles, it has been suggested, could be a substantial component of the cosmological missing mass, and would be expected to condense gravitationally, in particular, into galactic halos. (Wasserman 1986, p. 2071).

Both papers start out from the established effectiveness of neutrino detection techniques. They then immediately move on to the third premise: determining what would be required to make these neutrino detectors effective dark matter probes.

This also becomes the prime focus of Primack et al., and in their investigation of what properties dark matter would need to have, they immediately formulate arguments for the plausibility of these assumptions. The authors start out by listing various dark matter candidates. The list includes axions and light neutrinos,[13] but the main focus lies on WIMPs. Listed WIMP candidates include the lightest supersymmetric particle and the now-abandoned cosmion-proposal.[14] Setting up terrestrial detection experiments requires fine-grained assumptions about dark matter:

> In order to evaluate the proposed WIMP detection techniques, one must know the relevant cross sections [for the interaction of WIMPs with ordinary matter]. (Primack, Seckel, and Sadoulet 1988, p. 762)

The required detail is made plausible by appealing to possible cosmological scenarios that determine the dark matter abundance. The authors consider two:

1. The dark matter abundance today is determined by freeze-out as per the WIMP-miracle. In this scenario, dark matter is formed thermally in the early universe, just like baryonic matter, during Big Bang Nucleosynthesis. In order for the correct abundance of dark matter to be generated, stringent constraints on the allowed cross-section are placed.

[13]Light neutrinos were included as a clarifying contrast, rather than as genuine candidates, since they had already been rejected as plausible dark matter candidates based on their implications for theories of cosmic structure formation (Primack, Seckel, and Sadoulet 1988, p. 797).

[14]The cosmion was a theoretical particle that was supposed to resolve the solar neutrino problem, a discrepancy between the observed flux of neutrinos from the sun and the predicted rate based on the sun's luminosity. The solar neutrino problem was resolved by introducing neutrino oscillations. Cosmions were a rival solution: the weakly interacting cosmions supposedly increased the energy transport from the inner parts of the sun to the outer layers, thus reducing the central temperature required for the observed luminosity, and therefore the predicted neutrino flux. See Gelmini, Hall, and Lin 1987 for a detailed account of the cosmion proposal.

2. The dark matter abundance today is determined by a so-called “initial asymmmetry” (Primack, Seckel, and Sadoulet 1988, p. 763). This scenario posits the dark matter-abundance as an initial condition for the evolution of the universe, which means that “only a lower bound may be placed on the cross sections from requiring that annihilation be efficient enough to eliminate the minority species.” (Primack, Seckel, and Sadoulet 1988, p. 763)

The main purpose of both arguments is to investigate what assumptions about dark matter can plausibly be made such that direct detection experiments can be set up. Primack, Seckel, and Sadoulet (1988, p. 768) conclude from the cross section determination that “[t]he exciting possibility exists of detecting WIMPs in the laboratory”.

The remaining sections of the paper focus on details of the experimental setup. The discussion shifts from justifying the use of neutrino detection techniques for dark matter, to concerns about background mitigation and signal detection optimization. This discussion shows that making assumptions about the target system in order to justify why a particular method might detect it, does not exclude the usual experimental process of then minimizing systematic errors, or maximizing the signal-to-noise ratio.

2.4.4 Taking stock

Let me take stock. I have explored the particle physicists’ approach to dark matter experiments, and I have argued that these follow a method-driven (rather than a target-driven) logic for the justification of their effectiveness as dark matter searches. Specifically, physicists make additional assumptions about the particle properties of dark matter, most notably about its non-gravitational interactions (illustrated in figure 2.1). If dark matter has those properties, the various experiments could plausibly detect it.

What are the implications of relying on a method-driven logic? A first consequence is that the interpretation of the results is muddled. Suppose a direct detection experiment detects nuclear recoil by a dark matter particle from the galactic halo. That evidence could support two things: the fact that such a detected dark matter particle exists, as well as the assumptions that were required to justify the method choice in the first place. However, the detection can *only* provide evidence for the assumption that was required to get the evidence going in the first place, if it is plausible that no other unknown features of the target or the method could give rise to the signal. A second consequence is that the method-driven logic raises some problems for the appeal to common motivations for methodological pluralism in context of dark matter searches.

2.5 IMPLICATIONS FOR METHODOLOGICAL PLURALISM

In the previous section, I described two types of dark matter searches, each generating a separate line of evidence. This pursuit of multiple lines of evidence is common in scientific practice, and for good reason. One type of arguments concludes that it is *better* to have multiple lines of evidence than not. For one, increasing the empirical basis for an inductive inference is usually taken to strenghten the conclusion of that inference. Increasing the pool of evidence can also resolve local underdetermination problems and related issues in theory choice (see for example Laudan and Leplin 1991; Stanford 2017)—even if breaking so-called global underdetermination remains a lost cause. Another motivation comes from the literature on robustness.

Specifically for measurement results, Woodward (2006) defines "measurement robustness" as a concurrence of measurement results for the determination of the same quantity through different measurement procedures. A robust empirical result is a result that has been found to be invariant for at least a range of different measurement processes and conditions, and where a failure of invariance for certain other conditions or processes can be explained (Wimsatt 2007)[15]. Perrin's determination of Avogadro's number is a classic example of such a robust result. If different experimental procedures, with independent sources of systematic error all deliver the same value for a given parameter, the general argument goes, it is highly unlikely that the agreement is the result of different systematic errors all lining up, rather than that the result is tracking an actual physical effect. For example, in the neutron lifetime

[15]Wimsatt (2007) gives a general account of robustness analysis, which can apply to empirical results as well as features of theoretical models. Robustness analysis for modeling features has been discussed for example by Weisberg (2013), but it lies outside of the purpose of the current chapter.

example from section 2.3.1, the bottle- and beam-methods come with different sources of systematic error. Thus, if their outcomes agree, it is likely that the determined value of the neutron lifetime is accurate.[16]

Although this is not always explicitly stated, a crucial condition for measurement robustness is that the same parameter or quantity is being measured by the different experiments. This is where the definition of a target system comes in: the definition of the target system remains fixed under the employment of different methods. It provides, in other words, 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. While the common core does not guarantee that the same target is probed, I submit it is a necessary condition that this common core can be identified. Of course, the fact that the definition of the target system needs to remain fixed across various methods being applied to it, does not mean that it has to remain fixed over time. Rather, the common core can evolve, but that evolution needs to be shared across the different methods applied.

A second set of arguments concludes that there are contexts where methodological pluralism is not merely desirable insofar as possible, but in fact *necessary* to gain understanding of a complex target system and its workings in different environments. This idea of methodological integration is quite common in the life sciences. For example, O'Malley and Soyer (2012) describe how systems biology can be understood in terms of methodological integration, and O'Malley (2013) argues that various difficulties in phylogeny originate in a failure to apply multiple methods. Mitchell and Gronenborn (2017) use the development of protein science over the last five decades as an exemplar of how the hope for one experimental strategy to replace all others is futile. Kincaid (1990) (see also Kincaid 1997) formulates similar views with respect to both molecular biology and the social sciences. Finally, Currie (2018)

[16]Interestingly enough, the outcomes currently do not agree: the neutron lifetime measurements are resulting in two apparently inconsistent results, with the discordance currently at a 3.9σ significance (Pattie et al. 2018). The disagreement remains unexplained so far and has caused quite a stir, in part because it was unexpected given our knowledge of target system and methods predicting otherwise. In the case of protein-folding structure, NMR and X-ray crystallography also lead to diverging results, but Mitchell and Gronenborn (2017) show how this divergence can nonetheless lead to a more accurate representation of the protein folding structure.

describes the use of multiple methods as an integral aspect of developing an evidential basis for historical sciences.

Both motivations are at play in the dark matter searches. On the one hand, any result claiming to have found dark matter will only be accepted if multiple experiments belonging to multiple types of searches are able to find it. This is in part why the DAMA/Libra result, after several decades of claiming the sole positive result in dark matter searches, remains controversial in the larger physics world (see Castelvecchi 2018 for a recent update). On the other hand, different types of dark matter searches need to be employed to search as much area of parameter space as possible: production experiments are typically sensitive in a lower mass-range than direct detection experiments, where the deposited energy in the scattering interaction would be too small to pick up on. Direct detection and production experiments (along with the indirect detection experiments that were not discussed here) are therefore sometimes referred to as being "complementary"[17].

However, the use of the method-driven logic introduces a complication, both for measurement robustness, and for methodological pluralism more broadly in context of dark matter searches. To draw out this concern, let me recapitulate and bring together different points of the discussion so far. Recall that production and direct detection experiments require different assumptions about dark matter for the justification of their potential effectiveness. Moreover, the two types of searches are clearly two distinct methodologies to search for dark matter particles. In case of a positive detection, the two types of experiments could try to pick up the same signal, giving the basis for a typical measurement robustness argument. In the meantime, while such an uncontroversial positive result remains elusive, the two types can rule out complementary regions of parameter space. So where lies the problem for measurement robustness and methodological pluralism?

The situation is not quite as straightforward: complementarity comes at a cost. Direct detection experiments, unlike model-independent production experiments, can only search for one type of dark matter candidate at a time, with already specified coupling constants. Thus, direct detection experiments only constrain two free parameters: the interaction cross section

[17]For a discussion of complementarity in context of SUSY searches at the LHC, as well as its implications for perspectival realism, see Massimi 2018.

σ and the dark matter mass m_χ. Model-independent production experiments constrain four free parameters: the mass of the mediator particle $m_{mediator}$, the dark matter mass m_χ, the coupling at the standard model-vertex g_{ff} and the coupling at the dark matter vertex g_{DM} (see figure 2.1). The interaction cross section σ depends on all four of these parameters. To make results from production and direct detection experiments comparable—specifically, to translate production experiment results into constraints on m_χ σ-parameter space—requires making assumptions about two of the four free parameters, usually the coupling strengths g_{ff} and g_{DM}. These assumptions are again based on theoretical plausibility arguments, but they do pose a significant weakness for constraints from production experiments on specific dark matter models, since the translation of the production experiment results can be very sensitive to these assumed parameter values.[18]

2.6 CONCLUSION

I have proposed a new logic for justifying scientific method choice, which I have called a method-driven logic. This logic begins from a common core that describes a target system, but then looks at the method for guidance as to what further assumptions about a target system might need to be made in order for a method to be effective to discover new features of the target system. I submit that this logic is an interesting deviation from the more common target-driven logic.

The method-driven logic helps to understand why certain cosmological experiments have a chance at being successful, despite not being conducted on cosmological scales. It also raises an important puzzle for classic measurement robustness arguments in contexts where the method-driven logic is employed: due to the use of the method-driven logic, robustness arguments are not always readily available. This because of the different assumptions that

[18]This complementarity provides a different kind of data integration than the one discussed by Leonelli (2016) in context of plant science. In particular, there are no equivalents to iterative cross-species integration or inter-level integration, but rather a complementary probing of different but overlapping regions of parameter space. The main difficulty in this case is that different techniques are more effective for different regions of parameter space, but that the derivation of comparable constraints requires the above 'translation exercises'.

are required to start the experiment in the first place.

The puzzle for robustness and complementarity points towards some further puzzles: how should results, positive and negative, from method-driven experiments be interpreted in the first place? What do they provide evidence for (both with regards to the positive result, and the assumed feature of the target)? These questions only come to the forefront, once the crucial role of the method-driven logic is recognized.

3.0 Wishing for Failure? Interpreting Results from Method-Driven Experiments

The previous chapter introduced two structures of justification for the effectiveness of a particular experimental technique to generate evidence about a given target. I also briefly indicated that how an experiment is justified has implications for how results from that experiment can be interpreted. This chapter expands on those implications for method-driven experiments.

3.1 INTRODUCTION

The two structures of justification, which I call a 'method-driven' and a 'target-driven' logic answer the question: "which method(s) can reasonably be expected to be effective in producing evidence about a new feature of a given target T?" The two structures are the two extremes of a spectrum, where the spectrum itself is determined by how detailed and empirically confirmed the available model of the target is. The crucial difference between the two is that the target-driven logic relies on empirically confirmed theories or models of the target to select an appropriate method, while the method-driven logic adapts the theoretical description or model of the target to an available method. Unsurprisingly, the latter is particularly applicable in cases where the existing theory or model of the target is underdeveloped or not precise enough to appeal to in the justification of the method-choice. In that case, an entrenched experimental protocol may be repurposed to the new target, and the description of the target itself may be adapted so that the repurposed method may be optimally applied.

The use of either one of these logics has important implications for the interpretation of experimental results. In the previous chapter, I briefly highlighted the challenge of making experimental results from different method-driven experiments in dark matter searches comparable with one another. This chapter expands on that discussion. The main focus will be the following question:

> What is the data from a method-driven experiment evidence for? I.e., what conclusions can be drawn about a target T from empirical data that originate from a method-driven experiment—and specifically, does a positive (negative) result from a method-driven experiment provide evidence for (against) the assumptions about the target that were required to get the experiment going?

I will broaden the discussion here compared to the previous chapter by introducing several other examples of method-driven experiments.

I begin with a brief recap of the method-driven logic in the next section. I then consider two possible outcomes of a method-driven experiment: a positive result and a null-result (section 3.3). This discussion brings me to describing an apparent paradox of success: often, when a method-driven experiment seems to find a significant positive result, it is unclear what lessons about the original target can be drawn. But whenever a method-driven experiment fails to find a significant result, it is clear what one learns from that negative result about the target. Luckily, this ambivalence of a positive result is not definitive: in the fourth section, I discuss different possible strategies for dealing with the ambivalence of a positive result. Throughout the chapter, I will introduce several examples from scientific practice: cold nuclear fusion experiments, analog experiments in black hole research, and dark energy searches using atom interferometry, but I will also refer back to the dark matter particle searches introduced in the previous chapter.

3.2 METHOD-DRIVEN EXPERIMENTS: A BRIEF SUMMARY

The previous chapter introduced the target- and method-driven logics for justifying method-choice as two end-points on a spectrum. There, I argued that the method-driven logic is particularly useful in cases where the available theory of the target does not suffice to guide experimenters explicitly in the justification of their method-choice. Its overall structure is as follows:

P1: Given a target system T.
P2: Method 1 uses feature A of possible targets to uncover new feature X.

P2': Method 2 uses feature B of possible targets to uncover new feature X.
P2": Method 3 uses feature C of possible targets to uncover new feature X.
P3: If T has feature A, method 1 can be used to uncover new feature X.
P3': If T has feature B, method 2 can be used to uncover new feature X.
P3": If T has feature C, method 3 can be used to uncover new feature X.
P4: It is possible and plausible that T has features A and B, but it is either impossible or implausible (but possible) that T has feature C.

C: The preferred methods to uncover a potential new feature X will be methods 1 and 2, while method 3 remains out of consideration (for now).

Crucial in this structure are premises of types P3 and P4. In the remainder of this chapter, I will refer to premises of type P3 as *Assumption*, and to premises of type P4 as *Plausibility*.

Assumption introduces additional assumptions to the established theoretical description of the target. Premises of the *Assumption*-type thus allow scientists to justify choosing one method over another even if little is known about that target. *Plausibility* then ensure that these assumptions are indeed possible or plausible. 'Possible' indicates that the assumptions are not contradicting the independently empirically supported theories or models of the target (minimal though these may be), without being part of these theories or models. They are, in other words, still permitted. 'Plausible assumptions' are possible assumptions that are, in addition, further supported by plausibility arguments, for example, arguments by analogy, arguments appealing to simplicity, or arguments appealing to naturalness.[1]

I introduced the method-driven logic in the previous chapter by focusing specifically on contemporary dark matter research, and various particle physics experiments searching for a signature of assumed properties of the dark matter particle. For the purposes of the current chapter, it will be useful to get some more examples from scientific practice on the table as they become relevant. In the meantime, let me briefly reiterate the example of dark matter-searches.

In context of cosmology and astrophysics, dark matter was introduced as a non-baryonic type of matter. In context of particle physics, various searches for particle dark matter

[1] Dawid (2013)'s arguments for non-empirical confirmation (the no alternatives argument, the meta-inductive argument from past successful theories in the field, and the argument from unexpected explanatory connections) could also serve as plausibility arguments. Although Dawid maintains a categorical distinction between empirical and non-empirical confirmation, I remain wary of calling these arguments confirmatory at all. Nonetheless, I do agree these can serve as useful plausibility arguments in the sense defined here. I would like to thank Karim Thébault for this suggestion.

have been under way since the 1980s. In the previous chapter, I elaborated how both direct detection searches and production searches have been justified through a method-driven logic. For direct detection searches, the crucial *Assumption*-premise in the structure above was:

> **Assumption$_{\textbf{DirectDetection}}$**: If dark matter particles are weakly interacting, have a mass of the order of 100 GeV, and exist stably in the galactic halo, detectors similar to those used for neutrino detection can be used to search for signatures of dark matter particles existing in the halo of the Milky Way.

For model-independent production searches, the *Assumption*-premise was slightly different:

> **Assumption$_{\textbf{Production}}$**: If dark matter is made up of SUSY particles, or it couples to standard model particles through a mediator (invisible Higgs or otherwise), accelerator experiments can be used to search for signatures of dark matter particles or dark matter mediators.

Both of these experiments have been deemed as potentially effective to find (specific types of) dark matter particles, but, as discussed before, comparing their results is non-trivial.

With the general structure of a method-driven justification back on the table, let me now turn to the question raised in the introduction: what is the data from a method-driven experiment evidence for? And, specifically, does a positive (negative) result from a method-driven experiment provide evidence for (against) the new feature X, and the assumption that was required to get the experiment going?

3.3 THE APPARENT PARADOX OF SUCCESS

To get a full handle on these questions, I will consider two possible scenarios. Scenario 1 focuses on cases where a method-driven experiment gives a positive result, that is, where the experimental outcome seems to suggest that the target has the potential new feature X. Scenario 2 then examines cases where the method-driven experiment gives a negative result, that is, where the experimental outcome does not suggest that the target has a potential new feature X. Surprisingly, while it is unclear what a positive result reveals about the target, the conclusions from a null-result allow for strong conclusions about the target. This consequence of the use of a method-driven logic, I call the *apparent paradox of success* for method-driven experiments.

3.3.1 Scenario 1: a method-driven experiment gives a positive result

The first scenario I consider focuses on the positive case. Suppose a scientist justifies her choice of empirical method through a method-driven logic, and she concludes that the preferred method to uncover new feature X is method 1. She runs the experiment and finds a positive result. This result appears to provide evidence for the claim that 'method 1 finds that T has feature X'.

Recall, however, that the justification of the experiment relied on *Assumption*, i.e. the premise that 'if T has feature A, method 1 can find that T has feature X'. Thus, the positive result supports a slightly different claim. It supports the claim that 'method 1 finds evidence that the target T as described by existing theory *supplemented with an additional assumed feature A*, has feature X'. In other words, in order to justify the use of method 1 on the target T through a method-driven logic, the theoretical description of the target T has been modified to include an assumption about that target.

This modification has far-reaching consequences for what results from method-driven experiments can provide evidence for, as well as for how results of different experiments can be compared. It is no longer obvious that the positive result provides evidence for the claim that 'method 1 finds that T has feature X' because that conclusion relies on that additional assumption about the properties of the target. And it is equally not obvious that the positive result provides evidence for T having the assumed feature A. The question raised in the introduction (what is the data from a method-driven experiment evidence for?) thus encompasses two more specific questions.

1. Does the positive result provide evidence for 'method 1 finds that T has feature X'?
2. Does the positive result provide evidence for T having the assumed feature A?

I will use two further examples of method-driven experiments with a positive result to explore these questions: the cold fusion experiments from the 1980s (section 3.3.1.1) and the dumb hole experiments searching for Hawking radiation (section 3.3.1.2).

3.3.1.1 A primer on misapplying the method-driven logic

Example 1: The discovery and subsequent dismissal of cold fusion:[2] Nuclear fusion has long been sought after as a possible waste-free and inexhaustible solution to worldwide energy needs. Possible fusion reactions within current nuclear physics theories are the fusion of deuterium ^2H to produce large amounts of heat and fusion products (helium isotopes ^3He, tritium ^3H, hydrogen H and large quantities of neutrons n):

$$^2H + {}^2H \rightarrow {}^3He + H \tag{3.1}$$

$$2H + {}^2H \rightarrow {}^3He + n \tag{3.2}$$

In 1989, Martin Fleischmann and Stanley Pons released a press statement (University of Utah 1989), followed up by a publication in the *Journal for Electroanalytical Chemistry* (Fleischmann and Pons 1989), claiming that they had achieved nuclear fusion in a table-top experiment. The press release was euphoric, declaring "[t]he breakthrough means the world may someday rely on fusion for a clean, virtually inexhaustible source of energy", while also pointing out that "prior to the breakthrough research [...], imitating nature's fusion reactions in a laboratory has been extremely difficult and expensive" whereas Pons and Fleischmann's experiment was considered "equivalent to one in a freshman-level, college chemistry course".

The experiment consisted of the electrolysis of deuterium oxide (D_2O) with palladium electrodes over long time intervals.[3] The experiment was building on Fleischmann's experimental expertise. A renowned electrochemist, he had already done extensive work on electrolysis and, amongst other results, had found that prolonged electrolysis of D_2O could cause a build-up of deuterium on the palladium cathode (see for example Dandapani and Fleischmann 1972).

More specifically, Pons and Fleischmann reiterated that D_2O-electrolysis would lead deuterium to be highly compressed but lightly bound (and therefore highly mobile) on the

[2] The description of this example draws from Norton 2018; see also Storms 2007 for a biased but elaborate defense of the success of Fleischmann and Pons' cold fusion experiments.

[3] Although the paper makes no explicit mention of the duration of each experimental run, one side remark mentions "experiment times in excess of 120 hours" (Fleischmann and Pons 1989, p. 304).

palladium cathode. This suggested that "there must therefore be a significant number of close collisions and one can pose the question: would nuclear fusion of [deuterium] be feasible under these circumstances?" (Fleischmann and Pons 1989, p. 302).

In other words, Fleischmann and Pons used existing experimental protocols (electrolysis) to investigate whether deuterium can undergo the new and still mysterious process of cold fusion. In the argument structure from section 2, their method-driven justification required the following premise:

> **Assumption$_{\mathbf{CF}}$**: If deuterium can achieve high compression and mobility on a palladium cathode, deuterium can undergo cold fusion after prolonged electrolysis of D_2O.

Fleischmann and Pons considered the antecedent plausible based on Fleischmann's previous work. Their experiments supposedly showed that the consequent was true.

Despite initial excitement about the possibility of cold fusion, the results of Fleischmann and Pons were very quickly dismissed by the physics community. Norton (2018) lists four reasons that were given by the Energy Research Advisory Board (ERBA) as to why their experimental results did not sufficiently show that cold fusion had been achieved. First, despite some successful replications, there were many more failed replications. Second, calorimetric measurements were technologically difficult, and Pons and Fleischmann did not describe their measurement protocol in detail (except for an ominous remark that in one set of experimental conditions, large parts of their experimental apparatus melted due to excessive heat production (Fleischmann and Pons 1989, p. 305)). Third, the observations showed a discrepancy between the produced heat and the observed levels of fusion products. There were 10^9 less neutrons observed in the experiment than would be required for the produced heat to be the result of cold fusion reactions. Fleischmann and Pons (1989, p. 308) mention this in their paper, but they ascribed the discrepancy to some unknown nuclear process being involved rather than to a failure of nuclear fusion occurring. The fourth and final reason mentioned by the ERBA was that cold fusion should not be possible based on established theory: the electrostatic repulsion of nuclei should prevent them approaching closely enough to one another on the cathode for any nuclear fusion reaction to ignite.

All four reasons were important in the rejection of nuclear fusion, but the fourth reason is the most interesting for the current purposes. The method-driven justification for cold fusion experiments included the premise $Assumption_{CF}$, where the antecedent required that deuterium could achieve high compression and mobility, basically, to pack closely on the palladium cathode. However, nuclear physics theory prohibited this very possibility based on the electrostatic repulsion between atomic nuclei. This was reinforced by the failed replications, and provided sufficient reason for the ERBA and the larger physics community to ultimately reject the result as evidence for cold fusion.

The cold fusion case offers an interesting example of how the method-driven logic can be misapplied. Fleischmann and Pons made an assumption about the target (in this case deuterium) that was neither plausible nor possible according to the pre-existing theory of the target. In other words, their method-driven justification included premise $Assumption_{CF}$, but by doing so, made the fourth premise, *Possibility*, automatically false.

This example of a method-driven logic-gone-wrong reinforces the crucial role of the existing description of the target when applying the method-driven logic. Even though the description of the target might be thin, the additional assumptions about the target that the method-driven logic requires, need to be, at the very minimum, consistent with that description. Otherwise, the method-driven experiment cannot plausibly be claimed to probe the target it sets out to investigate, positive result or not.

3.3.1.2 A method-driven experiment gives a positive result. Now what?

I will set aside examples of misapplication for the remainder of this chapter. Suppose now that a scientist successfully and correctly uses a method-driven logic to justify her experiment, and the result of the experiment suggest evidence for the target having new feature X. What can she conclude about the original target T? A recent example from black hole research comes very close to this case.

Example 2: Dumb hole searches for Hawking radiation: Despite their clear mathematical description in general relativity, black hole physics remains notoriously devoid of empirical support. One example is Hawking radiation: Hawking predicted in the 1970s that black holes emit thermal radiation due to quantum effects near the black hole event horizon. Hawking radiation is expected to be very weak and therefore almost impossible to detect around black holes. Following an initial proposal from Unruh (1981), an alternative route has been proposed towards finding evidence for its existence: the use of so-called 'analog experiments' or 'dumb hole experiments'.

Their physical make-up can vary widely, from supersonic sound waves in a background fluid flow[4], to surface-waves in a water tank (Weinfurtner et al. 2010) or Bose-Einstein condensates (Steinhauer 2016). They all search for thermal radiation arising due to physics at the horizon, a phenomenon that is expected based on formal similarity between the mathematical descriptions of the dumb hole and the black hole, respectively. If such decisive evidence can be found—stronger than the current positive results—the hope is that these dumb hole experiments can be taken as evidence for the presence of Hawking radiation around black holes.

From this initial, rough description, dumb hole experiments might seem like an example of target-driven method choice. The theoretical model of black holes predicts that quantum processes at the horizon will give rise to Hawking radiation. This description, in combination with established isomorphisms with other physical models, implies that there would be similar processes at horizons for dumb holes as well, and that dumb holes can therefore be used to find Hawking radiation.

There is ground to doubt this target-driven reconstruction, however. Unruh (2014, p. 1) writes about Hawking's initial derivation that "while mathematically unimpeachable, they are nonsense physically". This because Hawking's derivations suffer from the trans-Planckian problem—in short, the fact that Hawking's derivation suggests the existence of radiation with absurdly large frequencies. As long as a complete theory of quantum gravity is lacking, the trans-Planckian problem causes significant worries for Hawking's derivation. Thus, it is not at all obvious that there is an empirically supported model of black holes to which a target-driven justification can appeal.

Instead, the justification of the effectiveness of dumb hole experiments to find evidence for Hawking radiation relies on various assumptions about the nature of black holes and their specific features that would give rise to Hawking radiation. Although it would lead me too far to fully elaborate the theoretical background behind the existing dumb hole

[4]These experiments are closest to Unruh's initial proposal, which he developed out of a pedagogical analogy. See Unruh 2008 for a full recounting of the pedagogical analogy, including the reader's imagining being a deaf physicist fish.

experiments, I can summarize Unruh (2014)'s defense of the effectiveness of dumb hole experiments (specifically, the experiments by Weinfurtner et al. 2010) as follows:

> **Assumption$_{\mathbf{BH}}$**: If black holes emit Hawking radiation such that the quantum emission (the Hawking process) is completely determined by the classical parameters of a linear physical system, and such that the essential features of quantum emission can be modelled in other physical systems than black holes, then dumb hole experiments can be used to find Hawking radiation.

The target in this case is black holes, the assumed feature is the universality of the quantum emission process under specific circumstances, the experimental protocol is the dumb hole experiment, and the new feature would be (a generalized version of) Hawking radiation.

So far, not all of the antecedent conditions have been met by Weinfurtner et al. (2010)'s experiment.[5] For example, quantum emission is supposed to happen spontaneously at the horizon, while the experiment only detected stimulated Hawking radiation. Moreover, the experiment only detected the thermal spectrum, none of the other defining properties of Hawking radiation. In fact, Unruh (2014) revealingly titled a follow-up paper "Has Hawking radiation been measured?" It is clear that these experiments have not been accepted as conclusive evidence for some quantum emission process taking place at black or white hole horizons.

Nonetheless, Unruh (2014, p. 8) does take the experimental result to give "strong support to the hypothesis that horizons, whether black hole, sonic or other will produce a quantum noise with a thermal spectrum, whose temperature is determined by the behaviour of the horizon". This bold claim raises the two questions from the beginning of this section: does the positive result provide evidence for the claim that dumb hole searches have found evidence for Hawking radiation? And does the positive result provide evidence for the universality of the quantum emission process under a specific set of circumstances?

There has been an ongoing debate about the confirmatory potential of analog experiment in the philosophical literature. On the one hand, Dardashti, Thébault, and Winsberg (2017) analyse dumb hole experiments as 'analog simulations' for black holes. They claim that if the isomorphism between modeling frameworks for black holes and dumb holes is robust,

[5] Steinhauer (2016) claims to also have achieved this success in a Bose-Einstein condensate analog system. For the scope of this chapter, I will not discuss his experiment any further.

dumb hole experiments can provide evidence for the universality of Hawking radiation and therefore for the existence of Hawking radiation from black holes. Crowther, Linnemann, and Wüthrich (2019) disagree: they claim that the dumb hole experiments beg the question when it comes to their applicability to black holes. Specifically, in order to confirm the presence of Hawking radiation around black holes from dumb hole results, it needs to be assumed that the modelling framework that gives rise to Hawking radiation is in fact applicable to black holes.

The debate reveals a broader issue with method-driven experiments and how to interpret a positive result. In order to conclude that a positive result of a method-driven experiment does, in fact, provide evidence for target T having feature X (in this case, that the detection of stimulated Hawking radiation in a dumb hole experiment provides evidence for black holes emitting Hawking radiation), the experimenter has to exclude alternative explanations for the experiment delivering a significant positive result. There are two parts to this.

First, potential sources of systematic error in the experiment need to be excluded. This is nothing out of the ordinary—any experiment requires this, whether method-driven or otherwise. For example, Belgiorno et al. (2010) made an earlier detection claim of Hawking radiation in an analog experiment. Their setup used laser pulses sent through a so-called non-linear material, that is, a material for which the refractive index changes under the influence of light. They detected the spontaneous emission of photons and claimed that this spontaneous emission was evidence for Hawking radiation at dumb hole horizons. The authors later retracted their identification of these photons with Hawking radiation, however, after strong criticism that the emitted light could be generated by other mechanisms (see for example Unruh and Schützhold 2012).

Second, and this is where the dumb hole-debate is situated, it needs to be established that the method-driven experiment, despite its justification relying on unconfirmed assumptions about the target, effectively probed the original target T. Again, in the case of Hawking radiation, the dumb hole experiments rely on the assumption that the quantum emission is completely determined by the classical parameters of a linear physical system, and that the essential features of quantum emission can be modelled in other physical systems than black holes. These are precisely the assumptions Unruh (2014, p. 9) attempts to argue in favour of, but he admits that further strengthening of the evidence is required.

This second point raises a new difficulty for method-driven experiments. In particular, it is true that all evidence claims require arguing that the empirical results are due to a real, physical effect, rather than due to an experimental error. However, the issue here is that the evidence has to be secured[6] against more than susceptibility to experimental errors alone. An additional argument needs to be made that the physical effect can be ascribed to (features of) the original target of interest.[7]

Now consider the second question raised at the beginning of section 3.1.1: does the positive outcome provide evidence for T having the assumed feature A? From the above discussion, it should be clear that this is not obvious from the positive result alone. But, as will become clear in section 3.4, the answer to both questions can be strengthened in case of positive outcomes.

Finally, the fact that a challenge arises for method-driven experiments that is different in kind to the challenges for target-driven experiments does not imply that evidence-claims from method-driven experiments will be less trustworthy than those from target-driven experiments. Generally, method-driven experiments are not less reliable compared to target-driven experiments; it is highly plausible that a target-driven experiment is susceptible to much worse systematics than certain method-driven experiments, for example. The securing of the evidence will be *different* for method-driven experiments; whether or not the security of the evidence itself is *worse* will vary on a case-by-case basis.

[6] I use Staley (2004)'s language of "security of evidence" here, which indicates the degree to which the claim "E is evidence for H" is susceptible to defeat from the failure of an auxiliary assumption. The 'security of evidence'-framework will play a more elaborate role in section 3.4.

[7] In addition to the dumb hole-example, I submit that this is also part of the issue with interpreting the controversial results from the DAMA/Libra experiment. The DAMA/Libra experiment (in combination with its predecessors) has consistently claimed to have detected dark matter signals since the 1990s. No other dark matter experiment has been able to detect a similar signal, however, and the results have generally been met with scepticism insofar as them constituting a genuine dark matter detection. Although I won't argue the case here, I believe that there is little doubt that DAMA/Libra has detected *something*, but there is significant doubt that their detection is one of dark matter particles.

3.3.2 Scenario 2: a method-driven experiment gives a negative result

What about the opposite scenario, where a method-driven experiment delivers a null-result for T having feature X? This scenario applies to various recent experiments in cosmology, including almost all current particle dark matter searches (discussed in the previous chapter) and to various dark energy experiments. Here, I only discuss one.

Example 3: Constraining dark energy with atom interferometry:[8] Dark energy, represented by the cosmological constant in the concordance model of cosmology, is the blanket term for whatever contribution to the total energy density of the universe is responsible for the accelerating expansion of the universe. It was introduced based on supernova redshift observations from 1998 and the later WMAP satellite measurements of the cosmic microwave background anisotropy power spectrum.

Currently, there is no accepted theory of the mechanism behind dark energy, although many proposals exist. A standard solution within quantum field theory (QFT) identifies dark energy with vacuum fluctuations, but this gives rise to the cosmological constant problem. The predicted value for the vacuum energy density based on QFT is around 120 orders of magnitude too large compared to the empirically determined value for the cosmological constant. Alternative proposals have to grapple with two further constraints: (1) preferably avoid introducing physics at energies far beyond the Planck scale, and (2) avoid mechanisms that produce a fifth force inconsistent with the strong limitations from terrestrial and solar system fifth force experiments. One possibility is to propose a scalar field whose properties vary with the environment that it is in. Fields of this kind have appropriately been dubbed 'chameleon fields': chameleon fields are screened in high-density environments (like the solar system) due to couplings between the scalar field and matter, and unscreened in low-density environments (like the near-vacuum of the universe).

Recently, atomic physicists have started to look for signatures of such a chameleon field through *light-pulse atom interferometry* experiments. The technique uses the same basic principle as light or gravitational wave interferometry, but applied to matter waves (see figure 3.1). Since the 1990s, atom interferometry has been used to measure various fundamental

[8] The description of this example is based on Hamilton et al. 2015; Burrage, Copeland, and Hinds 2015; Jaffe et al. 2017.

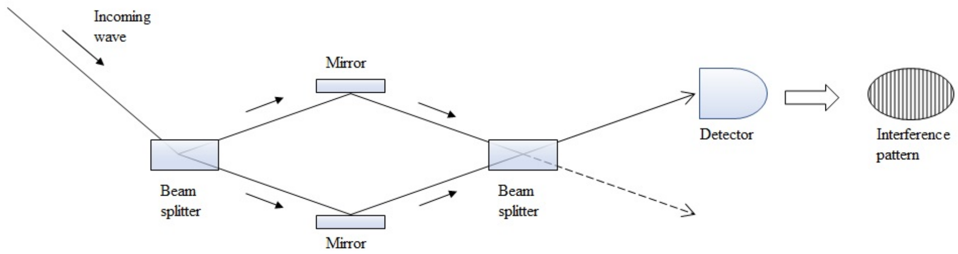

Figure 3.1: A schematic overview of the principle behind an interferometry experiment. An incoming wave is split, after which two waves follow two different paths. The waves along the two paths get reflected by mirrors, and recombined at a third point. The recombination gives rise to an interference pattern. Shifts in the interference pattern can hint at differences in path length, for example due to the stretching of spacetime by gravitational waves.

constants, e.g. the fine structure constant (for some recent results, see Bouchendira, Nez, and Biraben 2013; Parker et al. 2018).

Of interest here are a recent set of experiments that have repurposed light-pulse atom interferometry to dark energy research. In one experimental set-up (see also figures 3.2 and 3.3), caesium (Cs) atoms are launched into free fall in a scalar field gradient generated by a source mass suspended above the atoms. The Cs atoms will consecutively move through a beam splitter, a mirror, and a second beam splitter (in practice, these all consist of appropriately tuned laser pulses—hence the name of the experiments) with a separation time T between each beam splitter and mirror. Over the course of the experiment, the initial matter wave is split into two spatially separated[9] matter waves which are then reflected and recombined. At recombination, the phase difference between the two matter waves is determined based on the population ratio between possible quantum states $|a\rangle$ and $|b\rangle$, for a large number of atoms.

[9]The spatial separation is a result from a change in momentum that happens for each atom with probability 0.5. Two counter-propagating laser beams constitute the first laser pulse. The first laser beam will stimulate the atom to absorb a photon with momentum $\hbar k_1$. The atom then emits a second photon with momentum $\hbar k_2$ to the second, counter-propagating laser beam. The total momentum change for the atom is $\hbar k_{eff} = \hbar(k_1 + k_2)$.

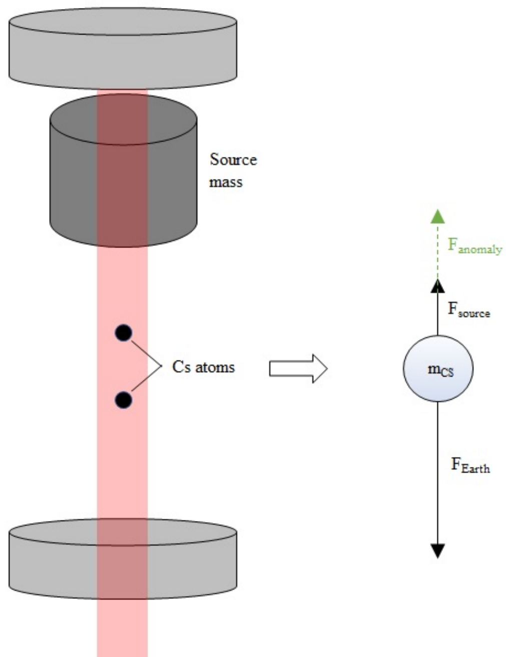

Figure 3.2: The basic set-up for a light-pulse atom interferometry experiment using Cs atoms with mass m_{Cs}. The atoms are launched into free fall, in the vicinity of a source mass mass m_{source}. While in free-fall they experience the gravitational force from the source mass (F_{source}), the gravitational force from the Earth (F_{Earth}), and potentially an anomalous fifth force ($F_{anomaly}$). Redrawn from Jaffe et al. 2017.

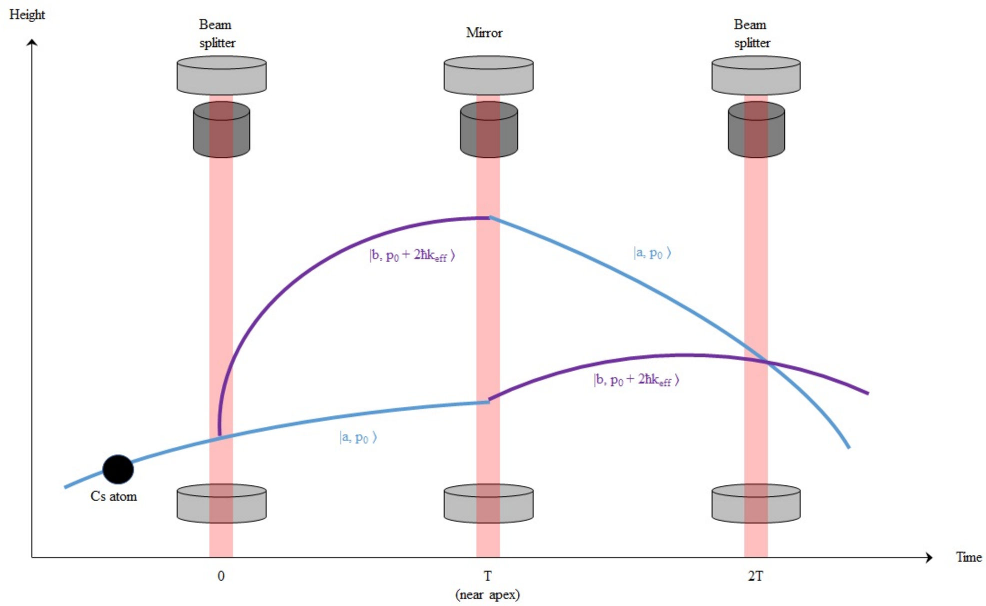

Figure 3.3: Detail of the three steps in a light-pulse atom interferometry experiment. The Cs atoms start in an initial quantum state $|a,\, p_0\rangle$. The first laser pulse's duration is tuned such that each Cs atom can make a transfer to another state $|b,\, p_0 + \hbar k_{eff}\rangle$ with probability 0.5. After time T, the two spatially separated matter waves are both 'reflected' by another laser pulse. This causes the atoms in state $|a,\, p_0\rangle$ to transfer to state $|b,\, p_0 + \hbar k_{eff}\rangle$, and vice versa. The spatial separation between the two waves remains. After another time interval T, the two matter waves are recombined by a third laser pulse. Redrawn from Jaffe et al. 2017.

To suppress the effects of the Earth's gravitational field, the experimenters opt to make a differential measurement between two possible set-ups: one where the source mass is located near the launched atoms, and one where it is far away from them. The acceleration arising from the source mass is then calculated as $a_{source} = a_{near} - a_{far}$. If there are no anomalous forces due to a chameleon field, this acceleration difference should be entirely due to differences in the gravitational force of the source mass. With the right value for the pulse separation time T, the source mass m_{source} of the gravitational field, and large enough samples of atoms, even small changes in a_{source} (and thus also the presence of a chameleon field) should be detectable (Jaffe et al. 2017, p. 938).

The light-pulse atom interferometry experiments are relevant for cosmology because, within the constraints from cosmology and fifth force experiments, chameleon fields are still a possible model for the fundamental physics giving rise to dark energy phenomenology. Banking on this possibility permits atomic physicists to justify that their experiments could be probing dark energy mechanisms. In the method-driven logic structure above, the crucial third premise becomes:

> **Assumption$_{\mathbf{DE}}$**: If dark energy is constituted by chameleon fields, then light-pulse atom interferometry experiments should be able to find a fifth force effect in the differential acceleration resulting from the chameleon field.

So far, no experiment has turned up a positive result (see also figure 3 from Hamilton et al. 2015). Burrage, Copeland, and Hinds (2015, p. 2) conclude that their experiments "show that there are already more constraints on chameleon scalar fields than previously thought".

The researchers draw strong conclusions about the exclusion of various regions of parameter space, based on their results.[10] This is symptomatic for what I consider a more general feature of method-driven experiments delivering a negative result. While the justification for method-driven experiments includes the *Adaptation*-premise ("if T has feature A, method 1 can be used to uncover new feature X"), the experimental outcome suggests that "method 1 does not find new feature X". From this, it follows that the outcome of the experiment

[10]The full parameter space encompasses all possible values for the free parameters in the chameleon field mechanism. Each set of parameter values is a possible specification of the chameleon field mechanism. By excluding regions of parameter space, a collection of possible specific chameleon scalar fields are being excluded.

supports (within the technical limitations of the experimental set-up) the conclusions that "T does not have feature A". Or, in other words, a negative result provides evidence against the assumption about the target that was required for the experiment's justification.

The opportunity to exclude possibilities reveals the power of method-driven experiments: they can narrow the space of possibilities significantly by excluding large regions of it. For example, the atom interferometry experiments have also failed to detect any signal of a chameleon field. This leads one group of scientists to conclude that "these fields are nearly ruled out, with only a one order of magnitude range left for the coupling strength" (Jaffe et al. 2017, pp. 939-940). In other words, atom interferometry has allowed scientists to conclude that chameleon fields most likely do not constitute dark energy, at least for a very large range of parameter values.[11]

But what about the new feature of the target that the experiment was supposed to detect but failed to? The fact that the experiment fails to detect that novel feature does not necessarily exclude T having feature X—that depends on whether or not having the assumed feature and using the specific experimental method are necessary conditions for detecting feature X or not. In practice, the assumed feature and the novel feature will often be related, for example when the assumed feature is a parameter range and the novel feature is a particular value within that range. In that case, the failure to detect X obviously also implies that T does not have feature X.

Before summarizing, let me make a quick note about the connection between this exclusion power of method-driven experiments and perspectival modeling. Massimi (2018) describes perspectival models as exploratory models that carve out the space of possibilities. She specifically introduces the use of models for physics beyond the standard model at the LHC, and argues that their crucial role is not one of modeling what is actually the case, but rather what could be the case. Accordingly, exploratory searches at the LHC that look for evidence in support of these models but fail to turn up a positive result still teach us something about what the world "cannot be like". The same is true for method-driven

[11] The same reasoning also occurs in dark matter searches. Dark matter direct detection searches have failed to turn up any positive WIMP result. This has put significant pressure on the WIMP hypothesis, and has even excluded the majority of obvious WIMP candidates, including most natural candidates for the lightest super-symmetric particle.

experiments delivering a negative result—they seem to fulfil Massimi's envisioned purpose for perspectival models.

3.3.3 The paradox of success for method-driven experiments

Let me take stock. I have considered two scenarios for the outcome of a method-driven experiment, and what claims these outcomes can provide evidence for. A negative result permits clear conclusions about the target, specifically about what features the target does not have. For a positive result, it is unclear what, if any, conclusion about the target can be drawn from the experimental outcome. This is what I call the *apparent paradox of success*: initial success turns out to be less conclusive with regards to the conclusions it warrants about the target, whereas initial failure is clear and conclusive. If one is looking for clear conclusions about the target, one seemingly needs to wish for failure.

Admittedly, things are not always as dire as I have presented them here for the positive results. It is possible that certain method-driven experiments have a particular relation between the assumed feature A and the detected feature X, such that, when the method-driven experiment does turn up a positive result, it can be considered convincing evidence for T having both the assumed feature A and the detected feature X. Some method-driven experiments might assume that a particular quantity of interest has a value that falls into a specified range (feature A). If the measurement result then is a particular value within that range (feature X), this result provides evidence for T having both features A and X. Those are realistic and ideal cases, but they are not universal; the dumb hole searches for Hawking radiation, for one, do not fit this ideal case.

The apparent paradox has some interesting implications. First, it showcases the power of a negative result. In case of method-driven experiments, negative outcomes can further constrain the space of possibilities for the target—a space of possibilities that might initially be extensive due to a very thin description of the target. Excluding possibilities can narrow down potential research avenues. For example, chameleon fields have almost entirely been rejected as a candidate explanation for dark energy. This is significant progress for a target about which very little is known. In the case of dark matter, the WIMP hypothesis has almost completely been excluded. Since no other candidate dark matter particle can be thermally created during Big Bang Nucleosynthesis, more attention is now turning towards baryogenesis, the process generating a matter/antimatter asymmetry in the early universe.

Second, the inconclusiveness of positive results of method-driven experiments has important consequences for integrating results of multiple methods being applied to the target, all through different instances of the method-driven logic: can the various positive results be used to strengthen one another? In other words, can they be used in a measurement robustness argument (see Woodward 2006 and the discussion in the previous chapter)? After all, each positive result relies on a different assumption about the target in the first place.

One might wonder at this point whether this concern of comparing positive results is unique to cases where different method-driven experiments give a positive result. Different method-driven experiments constrain parameter spaces for different assumptions about the original target, and it is not always the case that these parameter spaces are equivalent. This was already discussed at the end of the previous chapter, in context of dark matter searches. In order to compare the constraints, certain assumptions about the values of and relation between these parameter spaces need to be made. After all, different kinds of dark matter searches constrain different parameter spaces. What, then, is unique about the positive case?

The difference between the positive and the negative case is one between constraints on a full space of possibilities, and translating the constraints on the space of possibilities into constraints on one parameter space. The full space of possibilities can include different mathematical models, each associated with a specific parameter space. Negative results primarily constrain the full space of possibilities, which include these different parameter spaces. It is optional to then translate all negative results into a set of constraints on one specific parameter space. If this option is taken, the hurdles are more likely to be the same for the positive and the negative case, like in the case of complementary results for the WIMP parameter space.

Finally, making empirical results comparable (broadly construed) is not unique to the case of positive results from method-driven experiments. For example, Leonelli (2013) has discussed the issue of data integration in plant science, and specifically stressed the importance of different kinds of integration depending on (1) the question asked (investigating different levels in the same complex organism, or extending models from one organism to another?), and (2) the goal of the data integration (new knowledge, or new forms of interventions to improve human health?). Leonelli also discusses various sociological barriers to effective data integration like language, funding, resources, involvement of researchers in data curation (see also Leonelli 2016 for a detailed discussion of these topics). Although these are closely related to some of the issues discussed here, I submit that the use of a target- or a method-driven logic adds another determining factor as to how data integration is achieved.

3.4 STRENGTHENING POSITIVE RESULTS

There are strategies to alleviate the apparent paradox of success and to strengthen the positive conclusions of method-driven experiments for non-ideal cases. The various strategies divide into two categories: strengthening the evidence for T having the assumed feature A, or decreasing the dependence of the claim that T has feature X on the assumption that T has feature A. These two options parallel Staley (2004)'s options for securing evidence claims, i.e. for reducing the degree to which an evidence claim 'E is evidence for a hypothesis H' is sensitive to the failure of an auxiliary assumption of the experiment.

3.4.1 Find independent evidence for T having feature A

The first option is to find independent evidence that T has the assumed feature A (the antecedent of *Assumption*). This parallels Staley (2004)'s strategy of increasing the evidence for the auxiliary on which the evidence claim depends. In effect, this boils down to a post hoc-shift of a method- to a target-driven experiment: if there is independent empirical evidence for T having feature A, the problematic reliance on an unsupported assumption about the target can be alleviated.

This strategy might initially seem implausible in context of a method-driven logic. Why run a method-driven experiment in the first place, if it is possible to first find empirical support for the assumptions one is about to make about the target? Practical considerations could play a role. Sometimes it is easier or more cost-efficient to construct a wide-sweeping method-driven experiment first, for example. The atom interferometry experiments are much smaller and quicker than some of the current dark energy observational surveys like the Dark Energy Survey. Moreover, it makes use of existing expertise in atom interferometry rather than requiring the development and elaborate testing and optimization of new instruments.

Whether or not this first strategy is available depends on the nature of the assumption about the target. Consider again the dumb hole experiments. The assumption about the target was that the quantum emission is completely determined by the classical parameters of a linear physical system, and that the essential features of quantum emission can be modelled in other physical systems than black holes. This is summarized by Dardashti, Thébault, and Winsberg (2017) as the (qualified) universality of Hawking radiation. They argue that it would be possible to find further evidence for this universality by running more analog experiments in systems that share the required isomorphisms but have a different underlying microphysical structure. Thus, for certain kinds of assumptions about a target, finding independent evidence might be possible. Of course, this is the exact locus of the dispute with Crowther, Linnemann, and Wüthrich (2019), who argue that Dardashti et al.'s argument begs the question: it presupposes the black hole model one is trying to confirm through the analog experiments.

3.4.2 Decrease the dependence on T having feature A

Another option is to find additional evidence for T having the new feature X from different experiments, experiments that do not rely on the assumption that T has feature A. This reduces the reliance of the conclusion that T has feature X on the assumption that T has feature A, paralleling Staley (2004)'s suggestion to decrease the sensitivity of 'E is evidence for H' to the failure of the auxiliary assumption.

This additional evidence for T having feature X would, presumably, also come from a method-driven experiment. It would therefore suffer from the concern about comparing results from multiple methods that was raised before. Suppose method 1 finds a positive result, and a scientist wants to use this result to claim that T has feature X. She runs multiple different method-driven experiments and is lucky enough to keep finding positive results. Even though plenty of different method-driven experiments might all find a positive result, each of these results depends on a different assumption about the target, and this means that one can call into question whether the same target is being investigated by the various experiments.

Take, again, the dark matter particle searches from the previous chapter as an example. I ended the discussion there with a brief introduction to the complementarity of dark matter particle searches—specifically direct detection searches and production searches. The two types of experiments make different assumptions about dark matter: direct detection searches assume that dark matter consists of WIMPs, whereas model-independent production experiments assume that the dark sector couples to standard model particles through some mediator particle. Direct detection experiments, because they search for a specific type of dark matter, only constrain two free parameters: the interaction cross-section and the dark matter mass. Production-experiments constrain four: the mass of the mediator, the dark matter mass, the coupling at the standard model vertex and the coupling at the dark matter vertex (the interaction cross-section depends on all four of these). Suppose both a production experiment and a direct detection experiment find a positive signal. It is not at all obvious that these two positive results would provide empirical confirmation for the existence of the same type of particle or coupling.

There are at least two options for scientists to ensure that their different method-driven experiments all probe the same target T, despite the different assumptions they make about that target. The first is to formulate 'translation rules' between the different adaptations of the target. This is the current strategy in dark matter particle searches. If, once translated, the different positive results all suggest that T has feature X, the result that T has feature X seems less dependent on any single modification of the original target. Of course, these assumptions are themselves not necessarily based on experimental results, and must therefore always be handled with care. At the minimum, translation rules could allow consistency checks between results from different method-driven experiments.

A second option is to fully exploit the fact that the original target T is (close to) the only common denominator between the various experiments, for a larger number of method-driven experiments. This is what Dardashti et al. suggest about the Hawking radiation-searches, for example. The space of possibilities for a common explanation between the different positive results can be narrowed down by negative results (in line with scenario 2 in section 3.3.2). This would make the original target T having feature X the most plausible explanation for all of the individual positive and negative outcomes.

3.5 CONCLUSION

After introducing the method-driven logic in the previous chapter, this chapter delved further into the consequences of its application to specific experiments. The method-driven logic provided an alternative route to justifying method choice in situations where very little is known about the target. It remained an open question, however, how to interpret the results from these experiments. Here, I have shown that method-driven experiments can suffer from an apparent 'paradox of success': while a null-result allows for definitive conclusions about the target, a positive result is ambivalent, making failure almost seem like an appealing outcome. While method-driven experiments are powerful in excluding possibilities in light of negative results, their positive results remain ambivalent as to what they provide evidence for. However, certain strategies are available to strengthen these results: finding independent evidence for the assumptions about the target, or reducing the dependence of the result on the assumption. Although imperfect, they make wishing for failure no longer necessary.

4.0 Conflicts in Evidence

The previous chapters introduced how empirical data can be generated and interpreted in context of an underdescribed target. Here, I turn to conflicts between different lines of evidence, or, the 'problem of inconsistent evidence', specifically in a context where different disciplines are being integrated. Although contradictions between different lines of evidence prima facie pose a challenge for requirements of empirical adequacy, I show that the challenge sometimes can be overcome by ignoring part of the available empirical evidence.

4.1 INTRODUCTION

The plethora of proposals for dark matter and dark energy[1] (many of which originate from an experimental context where a method-driven logic is at work) comes with an interesting side-effect: an abundance of new lines of evidence suddenly becomes relevant to cosmology. For example, there are lines of evidence that potentially probe the fundamental nature of dark matter and dark energy. Even though these lines of evidence do not constitute observations of the universe, they nonetheless could be informative about that evolution of the universe.

The growth of the pool of evidence is not uniformly constructive. Sometimes, lines of evidence that supposedly confirm or falsify the same or related hypotheses about the target (here, the evolution of the universe) contradict one another. Stegenga (2009) describes this as one half of the 'hard problem of discordant evidence': the problem of *inconsistent evidence*.[2] Although Stegenga considers the problem of inconsistent evidence as a constraint on appeals to robustness, I take it to be a broader challenge for a general assumption in philosophy of science: that scientific theories should minimally be compatible with all available empirical

[1]For the purposes of the present chapter, I only focus on proposals that figure within the framework of ΛCDM, the standard model most cosmologists focus on developing further. I thus leave out of consideration alternative cosmologies like the ekpyrotic universe (first introduced by Khoury et al. 2001) or Roger Penrose's conformal cyclic cosmology (see for example Penrose 2014).

[2]The other half of the problem of discordant evidence is *incongruent evidence*, i.e. lines of evidence that seem incomparable with one another. The results of complementary dark matter particle searches, discussed in the previous two chapters, is an example of potentially incongruent evidence.

evidence. This assumption is shared by the constructive empiricist, according to whom the aim of science is delivering empirically adequate theories, and the realist, who can use empirical adequacy as a starting point for various arguments for (partial) realism. The problem of inconsistent evidence implies that it might be impossible to construct a consistent empirically adequate theory.

The goal of this chapter is to show that, at least sometimes, the problem of inconsistent evidence can be diffused by carefully considering what the pool of 'all available empirical evidence' should or should not include in light of the contradiction. In other words, I will argue that in some cases, not all evidence is equally relevant and I will introduce a relevance criterion (a solution Stegenga (2009) considers but dismisses out of hand). As will become clear below, that relevance criterion is a guideline with limited applicability, not an epistemological principle with an absolute certain guarantee for success. It does, however, demonstrate that the problem of inconsistent evidence is not an insurmountable challenge to the requirement of empirical adequacy.

To identify this relevance criterion, I first posit that, when modelling a complex target system, a distinction can be drawn between *unmediated* and *mediated* evidence (section 4.2). In section 4.3, I argue that, in case of contradiction between unmediated and mediated evidence, unmediated evidence can take priority over mediated evidence in two different ways: unmediated evidence trumps mediated evidence in confirming the model of the complex target, and it takes priority in guiding the construction of that model. Two case studies follow in sections 4.4 and 4.5, where I apply the abstract arguments to current research in dark matter (section 4.4) and to the history of the cosmological constant (section 4.5). The second case study draws out an important consequence of my view: two models of different target systems can be developed in parallel based on seemingly contradicting sources of evidence.

4.2 MEDIATED AND UNMEDIATED EVIDENCE

Consider a specific target system, i.e. a real-world system that is being modelled. In case of cosmology, the target system is the evolution of the universe on the largest scales. Often, the target system is different from the so-called object system, i.e. a system on which experiments can be performed. Different experiments can generate different lines of evidence about a given target. Not all empirical data will obviously contribute evidence for any possible target. For example, examining an onion peel through an optical microscope to reveal its cell structure will not obviously generate evidence relevant to modeling the functioning of the prefrontal cortex in a human brain. In order to constitute evidence for a model of a given target, there has to be some reason for the empirical data to be relevant for the specific target under consideration—whether it was generated through experiments on the target, or on an object system different from the target.

When modeling a complex target, there are two types of reasons why a particular piece of data can become relevant evidence for a model of a complex target system. The first reason is that a particular piece of data is believed to stand causally downstream to the target system for which the model is being developed. This notion of evidence is strongly reminiscent of the idea of "traces" in historical science, developed by Currie (2018, p. 70). Because of this causal relation, this piece of data contributes to delineating what the target system is, or it provides further evidence for the existing model of the target. Any successful model of the target system should account for the unmediated evidence, or at the very minimum be compatible with it. It is, in other words, constitutive of the explanandum, it constitutes the phenomenology the target gives rise to.

For example, in the case of ΛCDM, the target system is the evolution of the universe from an initial hot dense state up to today, approximately 13.8 billion years later. One of the crucial discoveries in the history of cosmology was that of a uniform radiation with a blackbody spectrum permeating the entire universe,[3] now known as the CMB. The observation of the CMB is a piece of unmediated evidence for any model reconstructing the evolution of the universe. The fact that steady-state cosmology could not explain the presence of the CMB, was part of the reason to dismiss it.

[3]The discovery of the CMB is often identified with Penzias and Wilson's observation of an isotropic background radiation in 1965. It is probably more historically accurate to consider the discovery as spanning several decades, from 1965 up until the first data results from the Cosmic Background Explorer (COBE) satellite in the 1990s. It was only with COBE that the CMB blackbody spectrum could be measured, for example. See Norton 2018; Ćirković and Perović 2016 for more details on the history of the CMB.

Not all evidence for ΛCDM falls under the category of unmediated evidence. Another reason why a piece of empirical data can become relevant evidence for a model of a target is that, despite being causally downstream from an object system different from the target system, the data is believed to be evidence for the model for other reasons. One example comes from the previous chapters: by employing a method-driven logic, new object systems are being experimented on to learn about properties of the target system. Other justifications draw on relevant similarities between object and target system, or on connections between scales. This evidence I call *mediated evidence*: it is evidence where the relevance is established through a different argument than its causal history.

For example, in context of cosmology, evidence from bubble chambers or particle accelerators provides unmediated evidence for the standard model of particle physics, and mediated evidence for Big Bang Nucleosynthesis (BBN), the theory of the formation of the lightest elements in the early universe.[4] Thanks to the discovery of the CMB, it is well-established that the early universe was in a hot dense state, similar to the types of states that can be recreated in particle accelerators. Thus, the types of processes that took place in the early universe can be described by the same theory that describes accelerator experiments, namely, the standard model of particle physics.

Although mediated evidence does not constitute the explanandum in the way that unmediated evidence does, mediated evidence can be crucially important when modeling complex targets in two ways. First, as the example of BBN illustrates, a model of a complex target is not conjured up out of thin air; it often incorporates different hypotheses and partial theoretical models that have been developed in a different scientific context. These partial models are applied to an entirely new target system, in this case the early universe. Mediated evidence then becomes important insofar as it provides independent support for the partial models that are being integrated. In other words, when available and not in contradiction with the unmediated evidence (unlike the two cases discussed below), mediated evidence ensures that the partial models and hypotheses feeding into the model of the complex target are on a secure footing.

Second, unmediated evidence can be gappy. Mediated evidence can therefore help solve

[4] See the next chapter for a more detailed discussion of BBN.

outstanding issues by providing additional empirical constraints, or even open up new research avenues. Consider again the appeal to nuclear physics in the modeling of BBN. The unmediated evidence consists of the abundances of the lightest elements in the universe, in combination with the CMB establishing the hot dense state of the universe. The mediated evidence from nuclear physics experiments here is crucial to provide further support for the modeling of BBN, e.g. for the determination of specific interaction cross-sections. Similarly, Currie (2018) describes how the existence of an enormous duck-billed platypus was reconstructed based on the discovery of one fossil molar, in combination with models and theories that had been developed in different areas of biology.

A couple of further remarks about the distinction introduced just now are in order. First, a piece of empirical data may provide both unmediated and mediated evidence, depending on the context in which it is being used. For example, data from accelerator experiments is unmediated evidence for the standard model of particle physics, but it is also mediated evidence for BBN. Similarly, while the CMB constitutes unmediated evidence for ΛCDM, it is now increasingly used as potentially mediated evidence for extensions to the standard model of particle physics (see the next chapter and Cyburt et al. 2016 for details). On a related note, over time the evolution of a field may result in a piece of previously mediated evidence becoming unmediated evidence, as the scope of the model widens. For example, while the formation and evolution of galaxy clusters is now also modelled by ΛCDM, this was not the case in the 1930s.

Second, a brief remark on the jargon introduced. I have chosen the language of unmediated and mediated evidence to indicate specifically whether the relevance of the evidence for the integrative model relies on the identity between the source of the evidence and the target system, or whether the relevance requires mediation by a different relevance argument. This of course does not mean that unmediated evidence does not require significant theoretical work to move from raw data to a piece of evidence; evidence claims are not bits of information that are free-floating in the world, ready for our taking.[5] That type of theoretical work will be required for both unmediated and mediated evidence.

This remark about jargon might lead one to think that the difference is primarily a

[5]For a clear elucidation of what theory-lading consists of in practice, see Boyd 2018.

difference in theory-lading between different lines of evidence. Bogen (2017) introduces three types of theory-lading: perceptual (differences in perceptual experience due to conceptual differences), semantic (the influence of theoretical commitments on descriptions of evidence), and salience (the influence of theoretical commitments on what evidence to collect). It should immediately be obvious why perceptual and semantic theory-lading are different from the unmediated/mediated-distinction. The distinction does track a qualitative difference in why a line of evidence is considered salient. Note, however, that this qualitative difference in what kinds of theoretical assumptions are loaded in a line of evidence does not imply a different degree to which a line of evidence is theory-laden or not.

Third, another seemingly related distinction is that between phenomenological and fundamental models. ΛCDM is taken to be a phenomenological model: it 'saves' observations of the universe at the largest scales, but it fails to explain what the fundamental physics is behind those observations. The phenomenological/fundamental distinction is a distinction between explanatory models, whereas the unmediated/mediated distinction is a distinction concerning what explanatory model a piece of evidence is supposed to be relevant for. They are therefore orthogonal to one another.

Finally, for the present purposes, I am assuming that both mediated and unmediated evidence are exactly that: *evidence*. I take this to mean that the procedures required to obtain evidence claims from initial detector data have been completed, and that, moreover, the resulting evidence claims have passed the standards of evidence that have been established by the relevant community. Thus, for mediated evidence from particle physics, for example, appropriate experimental protocol has been followed and the resulting statistical significance levels and systematic errors have been deemed acceptable by the particle physics community (standards different from what cosmologists might deem sufficient).

4.3 THE PRIORITY OF UNMEDIATED EVIDENCE

Unmediated and mediated evidence often both provide evidence for the model of the complex target system. Sometimes, however, unmediated and mediated evidence are in tension, or outright contradict one another with regards to specific hypotheses about the target: different measurements of the same parameter may disagree, causal structures may or may not be found, the target system may or may not have a particular property, etc. In Stegenga (2009)'s terms, the mediated and unmediated evidence may be inconsistent. Stegenga's concern is that "in the absence of a methodological metastandard, there is no obvious way to reconcile various kinds of inconsistent data" (654). What are scientists to do in the face of such contradiction?

As a starting point, consider the three ways scientists can move forward in the face of inconsistent evidence. The different ways forward depend on whether and where one is willing to locate the source of the contradiction: with the evidence, with the theoretical model, or with neither just yet. Let me examine the three options in more detail.

1. The first option is to *blame the evidence*. The scientist maintains the model of the complex system in its entirety, as well as the justification for the relevance of the evidence. She assumes that the error must lie with the evidence. If one assumes, as I have, that the evidence has passed the standards of the relevant communities, this error must be unknown. This first option of blaming the evidence thus boils down to banking on some unknown source of error. Without independent evidence for the source of this error, this can take decades to work out.
2. The second option, then, is to *blame the theoretical model of the complex system* itself. In this case, the scientist takes all the evidence seriously, as well as the various justifications for the relevance of the evidence, and therefore decides to completely change the model of the complex system such that the contradiction or tension is explained away. There is a concern that without additional evidence for some unaccounted for and specific physical effect causing the inconsistency, this option runs the risk of devolving into a series of ad hoc epicycles.
3. The third and final option is to *blame neither the evidence nor the model*. The scientist remains agnostic as to whether the source of the contradiction lies with the evidence or the model. However, in order to be able to make progress in developing the model in spite of the contradiction, the scientist limits the empirical scope of the model by declaring part of the available empirical evidence irrelevant for the model. Although this approach

is bold in one sense, in that it puts constraints on the requirement of empirical adequacy, it is conservative in the sense that it does not locate the source of the contradiction without sufficient reason to do so.

I have presented them here as separate options, but one might be concerned that the different options collapse into one another. For example, suppose a scientist changes the theoretical model such that the relevance of a piece of mediated evidence changes. Doesn't this de facto amount to a new source of systematic error? Although this might be the case in practice, the options track the locus of the explanation of the error. The first option claims that what presents itself as a systematic error is an unaccounted source of error associated with the measurement process, while the second associates the error with some unaccounted for physical effect. Similarly, one might argue that by declaring one line of evidence as irrelevant, the model of the target has de facto been changed. Hence, doesn't the third option collapse into the second option? The second option amounts to possibly changing the basic assumptions about how one should account for the unmediated evidence. The third option is potentially not as drastic: the relevance criterion I propose merely changes how a specific, local piece of mediated evidence fits in the integrative model. The theoretical change is located in different parts of the theoretical model, for the two options.

In my discussion of the three options, I have indicated that the first two are viable in cases where there is independent evidence for preferring that option. There is an asymmetry with the third option, which I described as more conservative in locating the source of the inconsistency. For the remainder of this chapter, I will focus specifically on cases where there is no additional reason to choose options one or two. Such cases seem to present the most pressing challenge of inconsistent evidence.

Stegenga (2009) indirectly recognizes the third option as a solution to the problem of discordant evidence, but dismisses this solution as merely shifting the problem of discordance to a problem of relevance. This because the third option only solves the problem of inconsistent evidence if one can find some reason to choose between the lines of evidence. How can one argue that one line of evidence is somehow more relevant to the model of the complex target than another? Here, I take up the challenge of providing a relevance criterion. I argue that unmediated evidence is more relevant for two reasons: prioritizing unmediated evidence is a more reliable strategy to develop an adequate[6] model of a complex target system, and unmediated evidence offers a more useful heuristic for developing a model of a complex system.

4.3.1 The argument from reliability

As the name suggests, the argument from reliability concludes that unmediated evidence is epistemically prior because relying on unmediated evidence can be a more reliable research strategy when (dis)confirming or constraining a model of a complex target system. It is more reliable because it does not require extrapolation from a different object system to the complex target system.

Integrative models are not conjured up out of thin air, but draw on pre-existing theories and models. These theories and models have been developed and justified in a different scientific context, using evidence from a different object system than the new target system, that is, using mediated evidence. The mediated evidence can be used to constrain the new model insofar as its source is relevantly connected to the target system. That extrapolation, however, is non-trivial. For example, Peebles, Page Jr., and Partridge (2009) explain how it was not obvious before the discovery of the CMB (and even for a while after!) that the early universe at some stage reached energy levels similar to those achieved in certain terrestrial particle physics experiments, and that, therefore, the ideas from particle physics could be used to explain the formation of the lightest elements in the early universe—even though Gamow's first proposals for BBN dated from the 1940s. This worry about the extrapolation does not arise to the same extent for unmediated evidence: unmediated evidence constitutes the explanandum itself. Unmediated evidence provides evidence for features that any model of the target system must, at minimum, be compatible with.

To put it differently, the argument from reliability draws on the difference in justification for why evidence can become relevant for a particular theoretical model. In the case of unmediated evidence, that justification is the causal relation to the target system. In the case

[6] I deliberately use "adequate" here because it can be interpreted in both a realist manner (where adequacy would indicate (approximate) truth) and an empiricist manner (where it would indicate empirical adequacy). The arguments below do not depend on the specific interpretation of adequacy.

of mediated evidence, the justification requires a different theoretical argument to establish relevance between the target system and the source system. That theoretical argument is fallible and a source of epistemic risk. This implies that relying on unmediated evidence is a more reliable strategy in situations where unmediated and mediated evidence contradict.

Two possible concerns about the argument from reliability merit discussion. First, as Currie (2018) indicates in the notion of traces, even evidence that is causally downstream from the target itself, requires some theoretical argument to establish that causal relation. Why can't it be equally plausible that the argument for the causal relation comes under pressure, rather than the relevance argument for the mediated evidence? Here, the argument from reliability assumes something that is the case for cosmology, but that might not always extend elsewhere. In case of cosmology, it is often the case that the theoretical argument for the causal relation draws from the model of the evolution of the universe itself. In other words, questioning the causal relation of the unmediated evidence would mean changing the model of the complex target system, an option that was dismissed earlier.

Second, the argument from reliability only applies in cases where a distinction can be drawn between evidence originating from the target system and evidence originating from other systems, *and* if it is assumed that both unmediated and mediated evidence are 'good quality' evidence that has been gathered through reliable methods and analyses, according to the standards of the relevant scientific communities. It is always possible that, in hindsight, there were additional sources of systematic error that needed to be taken into account. I will discuss a case below where it turned out that the unmediated, not the mediated, evidence was erroneous due to a missed source of systematic error. Such fallibilism does not give practical guidance, however.

4.3.2 The argument from heuristics

The second argument is the *argument from heuristics*. This argument focuses on how the scope and broad structure of a model of a complex target is developed in practice. Love (2008) introduces the notion of a so-called "problem agenda". When modeling a complex target, a problem agenda is a structuring of the complex question into more manageable parts. A problem agenda divides the research problem up in sub-problems and relations between the sub-problems, thus laying out a program for any modeling effort of a complex target.

Determining a problem agenda requires some existing knowledge about the complex problem that is addressed. The problem agenda is initially determined by the unmediated evidence, that is, by evidence for the different properties and behaviours that need to be explained by the model of the target system. Unmediated evidence sets the boundaries for future research: it delineates the space of possibilities (what is allowed?) and it determines the research questions (what needs to be explained?). This boundary-setting is the first step in determining what existing theories and models may apply to the complex target, and at what point these existing theories and models fall short and need to be extended. At later stages, uncovering further unmediated evidence can also change the problem agenda and lead to a reconsideration of where frameworks can and cannot apply.

For example, in the case of cosmology, the problem agenda addresses the overall reconstruction of the evolution of the universe from an initial hot dense state up until today. After the discovery of the CMB by Penzias and Wilson (1965), a main component of the problem agenda became accounting for the presence of this uniform source of radiation in the universe. Because the presence of the CMB revealed something about the expansion history of the universe, it became clear that Gamow's proposal was correct and that the formation of the lightest elements could indeed be situated in the early universe. Before the discovery of the CMB, the only available unmediated evidence was some crude estimates about the abundances of the lightest elements. Debates remained whether the lightest elements were formed in the early universe, as predicted by Big Bang cosmology, or whether they were continuously formed in stellar systems, as proposed by the rivalling steady-state cosmology. The CMB discovery provided sufficient unmediated evidence for the Big Bang scenario and for an early universe being hot and dense, which subsequently allowed for the further development of BBN.

Once a problem agenda has been established, it provides a first notion for how different mediated evidence may extrapolate to the integrative context, and how it might not. The problem agenda then guides the inclusion of the mediated evidence in the empirical basis of the theoretical model. Mediated evidence can provide further confirmation to certain hypotheses about the target, in addition to the already included unmediated evidence, or it can confirm theoretical conjectures about the target (for lack of unmediated evidence).

Determining how and when mediated evidence can inform a model of a complex system is not possible without having a problem agenda in place, even if that problem agenda is crude or contains many gaps. The problem agenda cannot be developed without referring to unmediated evidence to guide its construction. Thus, unmediated evidence can take priority in a second way: it serves as the primary heuristic when constructing an integrative model.

4.3.3 Taking stock

Let me take a step back. I have introduced a distinction between unmediated and mediated evidence. While both are important in scientific research on complex target systems, they sometimes are in contradiction. This raises the problem of inconsistent evidence. Resolving the inconsistencies could happen in various ways, as was discussed at the beginning of this section. Here, I have proposed two arguments that suggest restricting the scope of empirical adequacy for the model of the target and focusing solely on the unmediated evidence. Both guidelines go back to the fact that unmediated evidence constitutes the explanandum for the model of the complex system.

Again, this means that, by definition, the applicability of this relevance criterion is restricted to cases where the inconsistency is one between mediated and unmediated evidence (and where that contrast applies in the first place). Moreover, as I have indicated before, they only apply in cases where there is no independent evidence suggesting the influence of some unaccounted for experimental error or the presence of a new physical effect.

Another worry that may have come up at this point is that the priority of unmediated evidence may seem eerily close to the so-called 'materiality thesis' from the philosophical literature on simulations. In short, the materiality thesis claims that experiments on object systems that are made of 'the same stuff' as the target system are more epistemically powerful than, say, computer simulations, which have to rely on a formal instead of a material similarity. The materiality thesis has been argued against extensively, for example by Parker

(2009), Winsberg (2009), and Parke (2014).

The priority of unmediated evidence does not collapse into the materiality thesis, however. On the one hand, mediated and unmediated evidence can result from the same material, for example as is the case for the lightest element abundances (unmediated) and particle accelerator results (mediated evidence). On the other hand, the collection of unmediated evidence itself can result from a wide variety of 'stuff', ranging from galaxy clusters to the basic atomic elements.

It will be helpful to see how the previous arguments work out in practice and what their implications are. For this, I introduce two cases from cosmology, one contemporary and one historical. The tension between unmediated and mediated evidence differ slightly in each case, as does the type of justification for the mediated evidence. Nonetheless, it will become clear that the three options described at the beginning of this section descriptively capture the different responses to the tension in both cases. It will also become clear that prioritizing the unmediated evidence was a useful relevance criterion in resolving the tension between mediated and unmediated evidence, in absence of independent evidence for how to move forward when faced with that tension.

4.4 CONTEMPORARY DARK MATTER RESEARCH

The first case has already been discussed extensively in previous chapters: contemporary dark matter research. This research sits at the intersection between dark matter as a cosmological or astrophysical entity, and particle dark matter as a proposed extension to the standard model of particle physics.

As discussed in previous chapters, there is a wide range of cosmological and astrophysical evidence for the existence of some non-baryonic form of matter in the universe. One significant piece of evidence comes from the Bullet Cluster. This "direct empirical evidence for the existence of dark matter" (Clowe et al. 2006) consists of observations of a merging event between two galaxy clusters. Weak lensing maps reveal the gravitational effects of the Bullet Cluster, and thereby its overall matter distribution causing the lensing effects.

X-ray maps reveal the baryonic matter distribution in the Bullet Cluster. The two maps diverge, suggesting that there is additional non-baryonic matter dominating the total matter distribution. It is unclear whether and how dark matter's main rival hypothesis, Modified Newtonian Dynamics (commonly known as MOND, first proposed by Milgrom (1983)) can account for these observations.[7]

Although the different sources of evidence for the existence of dark matter function on different scales (from galaxies, over clusters, to the universe at large), they are all taken to exhibit phenomena that need to be (partially) explained by a model of the evolution of the universe. For example, structure formation requires the presence of additional matter to allow for the correct amount of gravitational collapse. Independently, galaxy rotation curves and the Bullet Cluster both provide evidence for the fact that galaxies are embedded in a dark matter-halo. In summary, an abundance of unmediated evidence supports the presence of dark matter in the universe.

From the perspective of particle physics, the situation is less promising. Several plausible dark matter candidates have been proposed, with most dark matter particle searches focusing on WIMPs. Recall from previous chapters that this is due to the so-called 'WIMP miracle': the predicted dark matter relic density for WIMP-dark matter matches the observationally determined dark matter relic density in the universe. Note that the WIMP miracle is partially based on unmediated evidence that determines the current best estimate for the relic density of dark matter in the universe (this includes the power spectrum of CMB anisotropies as well as Baryonic Acoustic Oscillations). The WIMP miracle is unique to WIMPs. Should dark matter be constituted by axions or sterile neutrinos, their abundance cannot be explained through thermal production. It either has been fine-tuned in the early universe to exactly the

[7] Although the Bullet Cluster is often considered the final nail in the coffin of MOND (not in the least by the scientists behind the discovery of the Bullet Cluster), MOND suffers from additional problems. Milgrom's initial, programmatic MOND proposal proved difficult to fit into a more realistic theoretical framework. First, it violated conservation of energy, momentum, and angular momentum. Second, it remained unclear how MOND could be embedded into general relativity. Subsequent proposals (e.g. AQUAL, RAQUAL, and Bekenstein's TeVeS) were meant to deal with these difficulties in a variety of ways. TeVeS was the most promising proposal for a long time, but it was recently excluded by LIGO's observations of the neutron star merger (Boran et al. 2018). Moreover, the search for independent evidential support for MOND remains inconclusive, partly because it has generally been difficult, if not impossible, to extract predictions from any of the MOND frameworks, especially on cluster scales. Finally, MOND fails to fill a crucial gap in the cosmological model of structure formation, something dark matter does succeed in.

amount required for large-scale structure formation in our universe, or it is an unexpected and unexplained result of baryogenesis. (Of course, the latter concern does not arise if dark matter is constituted by primordial black holes instead of a new non-baryonic type of particle.)

WIMP searches exploit the coupling of WIMPs to standard model particles through the weak interaction following the method-driven logic I have explored in a previous chapter. They broadly follow three strategies (see also figure 2.1): direct detection (see Freese, Lisanti, and Savage 2012 and Liu, Chen, and Ji 2017 for a review), production (see Buchmueller, Doglioni, and Wang 2017 for a review) and indirect detection (see Conrad and Reimer 2017 for a review).

So far, none of these approaches have led to a positive result, and not for a lack of trying.[8] Although large regions in parameter space remain to be explored, the latest generation of direct detection experiments is fast approaching the so-called 'neutrino floor' where WIMP interactions will no longer be distinguishable from neutrino interactions. As a result, the WIMP hypothesis has come under serious pressure. In a December 13, 2016 *ScienceNews* article, Rocky Kolb is quoted as saying that time is at least close to running out for WIMPs and that the next gerenation of direct detection experiments will most likely be conclusive about the fate of the WIMP hypothesis. Similar statements are made by Bertone and Hooper (2016). This raises an interesting tension in contemporary dark matter research. While there is an abundance of unmediated evidence for the existence of dark matter from cosmology and astrophysics, there is no positive mediated evidence on the particle physics side (be it for WIMPs or for other dark matter particle candidates).

Different responses have been given to the tension. Some, mostly proponents of MOND and other alternative theories of gravity, hope that the lack of positive results in WIMP searches will lead to an abandoning of the dark matter idea altogether (see for example McGaugh 2015 and Narlikar 2018). The majority of the cosmology community takes a much less drastic approach: they refuse to give up on the abundance of cosmological and astrophysical evidence for the existence of dark matter. Rather, they seriously reconsider the importance of the WIMP miracle. These two responses broadly correspond to a complete

[8] I am deliberately leaving aside the controversial and so far unreplicated DAMA/Libra result.

revision of ΛCDM on the one hand, and revising the correctness of the relevance argument for a specific line of mediated evidence on the other. Given the number and variety of experiments, there is very little reason to believe that the lack of WIMP detection is due to some unknown error in the experimental results. That option has therefore been left out of consideration in the remainder of the discussion.

The two guidelines introduced above, help explain why revising ΛCDM in light of the failure of the WIMP hypothesis is not recommended. First consider the argument from reliability. Although the available cosmological and astrophysical evidence puts some constraints on the fundamental dark matter particle nature, it leaves open a space of possibilities for dark matter particle candidates that extends far beyond WIMPs. Taking the failure of WIMP searches as evidence against the dark matter hypothesis in general, assumes that WIMPs are the only possible dark matter candidate. There is no empirical evidence supporting this assumption, however (recall from chapter 2.0 that the WIMP miracle makes WIMPs a plausible dark matter candidate, but it is entirely possible for particle dark matter to be created during an earlier stage of the universe, for example, or for dark matter to be constituted by primordial black holes!). By refraining from rejecting the dark matter hypothesis, despite the lack of evidence for the existence of WIMPs (and, indeed, the exclusion of large swaths of WIMP parameter space), defenders of the concordance model of cosmology remain well within the constraints posed by the available unmediated evidence.[9]

This is different for certain defenders of MOND who reject dark matter in light of the failure to find WIMPs. Instead of including all available unmediated cosmological evidence as relevant evidence for cosmological theorizing, they restrict the relevant evidence to galactic scales (at which MOND operates), as well as the negative evidence from failed WIMP searches. What is not included as relevant evidence, according to this approach, is the evidence from the CMB or from the Bullet Cluster. In other words, this approach fails to account for part of the unmediated evidence, and, as a result, completely revises the cosmological concordance model. Given the general argument from reliability, this is an epistemically inferior strategy compared to the previous one.

[9]This is also in line with what I argued in the previous chapter. The negative results from WIMP searches have shown what dark matter can not be constituted by. This is different than claiming they provide evidence for dark matter not existing altogether.

Second, this case also shows how the argument from heuristics suggests that a new relevance arguments can be constructed to move dark matter reserach forward. Foregoing the WIMP miracle (the majority approach) means that a new justification needs to be constructed to justify the relevance of particle physics to the dark matter question and for various new experiments to be constructed following a method-driven logic. Unsurprisingly, this is no easy task, as it is not obvious what alternative dark matter candidate those experiments should look for. While most alternative candidates fit within the current cosmological constraints, all come with the disadvantage that their abundance cannot be causally explained in context of BBN.

However, as the argument from heuristics suggests, there is progress to be made even without a new relevance argument immediately on the table. There is ongoing research on cluster phenomenology, structure formation (for example with regards to the small-scale challenges and what these could reveal about the nature of dark matter[10]), and specific events like the Bullet Cluster, that is specifically focused on further constraining the space of possible dark matter candidates. This approach is guided by new or improved unmediated evidence and without a complete overhaul of the current concordance model—as suggested by the argument from heuristics.

4.5 COSMOLOGICAL TIMESCALES IN THE 1930S

The dark matter case illustrated how different responses to inconsistent evidence can be classified according to the broad categories introduced at the beginning of section 3. It also illustrated how prioritizing unmediated evidence is the preferred relevance criterion when restricting the evidential basis for the model of a complex target system. As a second example, I turn towards the history of cosmology and the problem of mismatching timescales in the 1930s. While in the case of dark matter, the inconsistency is easily resolved because

[10] The small-scale challenges, which include the missing satellites problem, the too big to fail problem and the cusp/core problem are an area of ongoing research. Although unresolved, there is at least some indication that these problems may be due to simulation assumptions, rather than some fundamental new property of dark matter. See Bullock and Boylan-Kolchin 2017 for a review of the problems as well as possible resolutions.

of the large space of possibilities beyond WIMPs, the current case provides an example of an inconsistency that seems, at first sight, much harder to resolve. Still, I will show that the priority of unmediated evidence applied here as well.

4.5.1 Mismatching timescales

The episode of mismatching timescales goes back to the early days of relativistic cosmology in the first half of the 1930s. Hubble (1929)'s distance measurements and observations of redshifts of Cepheid variables had led to the acceptance of the first physical model of an expanding universe (Lemaître 1927; Eddington 1930; Lemaître 1931). One important detail about the model is that, like Einstein (1917)'s earlier static model, this expanding universe model included a non-zero cosmological constant.

Although they were a milestone for relativistic cosmology, Hubble's redshift observations introduced a new problem: they suggested an expansion age[11] of the same order of magnitude as geological timescales from radioactive dating methods, and of a smaller order of magnitude than the accepted estimates for the timescales of stellar evolution. The conflict transformed into an outright contradiction by the common assumption that stellar evolution could not have commenced until after the onset of the expansion of the universe (an assumption based on Eddington (1930)'s work on the instability of the Einstein universe). In other words, the unmediated evidence for the model of the evolution of the universe, in this case Hubble's estimate for the age of the universe, contradicted the mediated evidence from stellar evolution and geology.

The conflicting timescales were widely recognized as problematic (see Eddington 1930; Tolman 1934, and personal correspondence between Einstein and Lemaître[12]). At the centennial meeting of the British Association for the Advancement of Science (BAAS), in late 1931 (and reported in British Association for the Advancement of Science 1932), several lead-

[11]Hubble's redshift observations lead to the famous distance-velocity relation, or the Hubble-Lemaître law: $v = Hd$. In this equation, v is the velocity of the astrophysical object (which can be determined based on its redshift), d is the distance from the observer to the object, and H is the Hubble parameter (indicated as H_0 for the universe today). $\frac{1}{H_0}$ gives a rough estimate for how long the universe has been expanding, and is therefore commonly used as an estimate for the age of the universe.

[12]Available at the *Archives Georges Lemaître*, Université Catholique de Louvain, Louvain-la-Neuve, Belgique.

ing cosmologists gathered to discuss the evolution of the universe. Many, including James Jeans, Arthur Eddington, Willem de Sitter and Georges Lemaître, explicitly addressed the mismatching timescales. de Sitter, for example, said:

> Now if we adopt this theory of the expanding universe, it is very tempting to seek a connection between this expansion and the evolution of the material bodies constituting the universe, and to identify the beginning of the expansion with the beginning of that evolution. But the time elapsed since the beginning of the expansion is only a few thousand million years—an interval that we have learned to consider as very short from the evolutionary point of view. There is no escape from this. (British Association for the Advancement of Science 1932, p. 585)

Here, de Sitter draws attention to the tension between the age of the expansion of the universe and the age of the oldest stars. Moreover, he suggests that the connection between the expanding universe and stellar evolution was non-trivial. This seems to fit well with the distinction between unmediated and mediated evidence.

Two types of responses to the contradiction came into play at the meeting. On the one hand, Lemaître took the contradicting timescales as sufficient reason to change the cosmological model that he had first proposed five years earlier. Lemaître proposed an entirely new expanding universe model which included a 'fire-works beginning of the universe',[13] an initial phase of expansion, a period of stagnation (similar to a static Einstein universe) and a second phase of expansion. This 'coasting model' of the expanding universe solved the timescale problem by explaining away the discrepancy (following the second option from section 4.3). Hubble's redshift measurements only provided an estimate for the duration of the second expansion phase, while the evolution of stellar objects could already commence during the first expansion phase and the coasting phase.

This model gained little traction. This was partly because it still included a non-zero cosmological constant Λ, which was required for the stagnation phase. However, Einstein and Sitter (1932) had developed an expanding universe-model without Λ around the same time that the timescale discrepancy emerged. With this new model of the expanding universe, a large majority of cosmologists dropped the cosmological constant. Lemaître refused[14] because of the conflicting timescales, a tension that the Einstein-de Sitter model in itself

[13]He also referred to this fire-works beginning as a "primeval atom", see for example Lemaître 1931.
[14]See Lemaître 1949 for a passionate defense of Λ.

could not resolve.

But even before Einstein and de Sitter published their expanding universe model in 1932, few of Lemaître's peers were keen to accept Lemaître's changes to the cosmological model. The continuation of de Sitter's address at the BAAS meeting is particularly clear in this regard:

> I do not think it will ever be found possible to reconcile the two time scales. We thus, however reluctantly, come to the conclusion that the expansion of the universe on the one hand, and the evolution of stellar systems and stars on the other hand, are two different processes, taking place side by side, but without any apparent connection between them. (British Association for the Advancement of Science 1932, p. 585)

de Sitter makes a very strong claim about the future of cosmology as a field. Even if interpreted in a slightly weakened version, about the state of cosmology at the time, the claim is informative. de Sitter suggests that he would rather accept the disconnect between the evolution of the universe and that of its components, than to immediately accept radical changes to the model of the evolution of the universe.

This general attitude towards the timescale problem prevailed after 1932 and was shared by many of de Sitter's colleagues. Tolman, for example, explored the possibilities of a big bounce resolving the timescale discrepancy (see for example Tolman 1934, as well as de Sitter 1933). This model did not require changing the model of the expansion phase, but rather explored the initial singularity such a model might suggest. Einstein, from his end, claimed that the time scale discrepancy only arose because of the idealizations in the expanding universe model. He believed that taking into account inhomogeneities could most likely resolve the problem (he includes this defence in personal correspondence between Einstein and Lemaître[15]).

The main upshot here is that, unlike Lemaître, his colleagues were all weary of changing the core of the cosmological model. Rather, they put into question the specific justification for the mediated evidence being relevant for the evolution of the universe, in this case that stellar evolution only commences *after* the homogeneous and isotropic universe has started to expand.

[15] Available at the *Archives Georges Lemaître*, Université Catholique de Louvain, Louvain-la-Neuve, Belgique.

4.5.2 Why restricting the empirical scope to the redshift observations was justified

Lemaître's colleagues had good reason not to accept the reintroduction of Λ or Lemaître's primeval atom, when framed in terms of the priority of unmediated evidence. First, as far as the argument from reliability is concerned, de Sitter's restricting of the empirical scope of the cosmological model was a reliable strategy. Hubble's redshift observations were something any cosmological model would have to explain—something further elucidated by the fact that Hubble's data were solely responsible for a breakthrough in the Einstein/de Sitter-debate that started in 1917. For the mediated evidence from geological or stellar timescales, however, no model existed for planet or star formation in the expanding universe. It was therefore unclear how exactly stellar evolution provided a precise estimate for the age of the evolution of the universe.

Second, consider the argument from heuristics. Relativistic cosmology was very much in flux. Although a uniform expansion of the universe was widely accepted, the proposed models varied widely in what they posited about the aforementioned initial state of the universe, or its matter content (Tolman (1934)'s textbook provides a good overview), all of which would be compatible with the redshift observations. There were no further observations suggesting how to differentiate between them. The problem agenda at the time was crude, and it was yet unclear where geology and theories of stellar evolution would all fit on the cosmological timeline. The redshift observations provided some guidance to how the evolution could be modeled, but more evidence was needed in order to add stellar and geological timescales in that evolution.

Two further lessons can be drawn from this episode. First, the priority of unmediated evidence provides the required guidance to move forward in spite of inconsistent evidence. It does not imply that unmediated evidence cannot be fallible, just like mediated evidence. In the 1950s, it turned out that a mistake in Hubble's distance measurements laid at the basis of the timescale discrepancy. Baade (1956) and Sandage (1958)'s corrections to the distance measurements for Cepheid variables ultimately resolved the timescale discrepancy.[16]

[16] A new version of the timescale discrepancy emerged again in the 1990s, when the age of globular star clusters appeared to be older than the age of the universe (see Bahcall 1995; Bahcall 1996). This was resolved

These corrections were made to the unmediated evidence, independently from the mediated evidence.

Second, the episode illustrates an important consequence of the distinction between mediated and unmediated evidence: situations may arise where it is reasonable for different scientific communities to work in parallel with one another based on contradicting evidence. This was clearly the case for geologists and cosmologists between 1930 and 1958. Each community focused on accounting for its own unmediated evidence first, and independently continued to improve on those lines of evidence.

4.5.3 The reintroduction of Λ

It is useful to contrast the debate between Lemaître and his peers with the reintroduction of the cosmological constant in 1998. Riess et al. (1998) and Perlmutter et al. (1999) independently reported on anomalous observations of Type Ia Supernovae (SNe Ia) redshifts.[17] These redshifts did not contradict, but complemented previous measurements by examining supernovae at higher redshifts than before.

In 1998, the model of the universe was a matter-only universe including baryonic and dark matter. Since the model predicted a very small deceleration of the expansion of the universe due to gravitational attraction, it predicted an almost linear relationship between the SNe Ia luminosity and their redshifts at lower redshifts (i.e. closer to the earth). These predictions fit well with observations.

The surprise in the 1998 observations was that the observations were probing SNe Ia at higher redshifts, higher than had ever been observed, in an attempt to measure the deceleration rate of the expanding universe. The observations revealed that the luminosity of those high-redshift SNe was lower compared to what was expected based on the matter-only model of the universe. This diminished luminosity provided a new type of unmediated evidence for the evolution of the universe. In particular, the diminished luminosity revealed

by the discovery of the accelerating expansion of the universe. I would like to thank Arthur Kosowsky for this point.

[17]Type Ia Supernovae result from a binary system of a white dwarf and a massive star. SNe Ia supposedly have extremely regular absolute luminosities thanks to uniformity in the critical mass at which the white dwarf can no longer support additional mass accretion and explodes. They serve as so-called 'standard-candles' for astronomical distance measurements.

that the expansion of the universe was accelerating, rather than decelerating, an idea that was reinforced by the WMAP measurements of the power spectrum of the CMB anisotropies. Perlmutter, Riess and Schmidt shared the 2011 Nobel Prize in Physics for their discovery of the accelerating expansion of the universe, and the amalgamation of unmediated evidence for the accelerating expansion warranted a radical change of the general cosmological model at the time.

The crucial difference between this episode and the previous one is that the reintroduction of Λ was motivated solely by the unmediated evidence, and that the new line of evidence extended the existing evidence, without necessarily contradicting it. This partially explains why the discovery almost immediately changed the concordance model of cosmology (only 13 years elapsed between the initial publication and their Nobel Prize win). Moreover, their discovery came at a time when the model of the evolution of the universe, as well as of the evolution of supernovae and the formation of large-scale structure was already in a much more developed stage, and it was clear that only modifications to the Friedman equations (within general relativity) would be able to account for the new set of observations.

4.6 CONCLUSION

I began by introducing an apparent threat to the requirement of empirical adequacy for scientific models: the problem of inconsistent evidence. Here, I discussed this threat particularly in context of models of complex targets, where a distinction between unmediated and mediated evidence can be drawn. I assessed different possible responses and proposed a relevance criterion, which amounts to a rethinking or refining of the requirement of empirical adequacy: in case of conflict, the most basic requirement of empirical adequacy is limited to unmediated evidence.

This relevance criterion also allows for a more precise articulation of what guides the modeling of a complex system and what the pitfalls for integrating various disciplines can be. First, the modeling effort is guided by the target system, and in particular by the unmediated evidence originating from that target system. It is this unmediated evidence that determines the problem agenda and that constrains the questions and answers given about the target. This means that delineating the target system, as well as how the target system changes over time, is of crucial importance.

Finally, the relevance criterion only applies in cases where the difference between mediated and unmediated evidence applies. The problem of inconsistent evidence is not restricted to those cases, however. Nonetheless, this does not obviously imply that the requirement of empirical adequacy is a lost cause. Although a comprehensive overview would lead me too far, I submit that in other cases of inconsistent evidence than the ones discussed here, there are ways to move forward. For example, Mitchell and Gronenborn (2017) argue that inconsistent evidence can allow for mutual corrections between results from different experimental techniques. Alternatively, Bernal and Peacock (2018) suggest that seemingly inconsistent parameter determinations can nonetheless be amalgamated into a more conservative estimate for that parameter. These two examples provide some reason to believe that the problem of inconsistent evidence can quite often be solved.

5.0 Coupling Back from Cosmology to Particle Physics: A New Cautionary Tale

Previous chapters have all introduced epistemic issues that arise when particle physics contributes experimental techniques and lines of evidence to answer questions in cosmology. The fact that cosmology provides a new research area for high-energy physics to contribute to, also means that high-energy physicists have started to use cosmology to break new grounds on remaining questions about the standard model of particle physics (and its extensions). This chapter first introduces the methodological shift associated with this turn to the early universe, and then locates the epistemic shift associated with that methodological shift.

5.1 INTRODUCTION

The European Strategy for Particle Physics is being updated between 2018 and 2020. This collective decision-making process was launched in 2005 and updated a first time in 2012. Both times, the Large Hadron Collider (LHC) was considered the top priority for the future of particle physics. The current update is focused on post-LHC facilities, since the LHC's run is expected to end in 2035, and setting up new experiments will require significant time and resources. Some of the proposals include the International Linear Collider (ILC) in Japan, the Compact Linear Collider (CLIC) and the Future Circular Collider (FCC) at CERN's existing site, or the Circular Electron Proton Collider (CEPC) in China. Each proposed experiment would surpass the LHC's reach, either by colliding electrons and positrons at similar energy levels to the LHC (ILC, CLIC), or by colliding hadrons at much higher energy levels (FCC), or both (FCC, CEPC).

The urgency of these experiments is usually defended by appealing to the incompleteness of the current Standard Model (SM) of particle physics. For example, in a design report for CLIC, it is claimed that "[new particles interacting with the SM with electroweak-sized couplings] are expected because many of the shortcomings of the SM are inherent to the electroweak sector of the theory" (CLIC and CLICdp 2018, p. 3). The shortcomings of the SM include the lack of a convincing dark matter candidate, no mechanism for charge or charge/parity-violation (necessary to explain the baryon/anti-baryon asymmetry in the universe), or the naturalness problem.

Building a new accelerator that can reach higher energy-levels or lower background-noise thus seems like an obvious way forward, especially since it is in line with particle physics recent history of success.[1] The 2012-discovery of the Higgs boson is arguably the crowning achievement of this strategy so far. However, building a post-LHC accelerator also comes with significant practical challenges. For one, each of the options currently on the table will cost billions of dollars in construction and operating costs. Moreover, it is not clear that the cost will pay off, since the space of possibilities for physics Beyond the Standard Model (BSM) is much larger than any single experiment can probe. It is therefore not all that surprising that even the ILC, which is furthest along in its development, has yet to receive full commitment from the Japanese government to host the experiment (Banks 2019).

In light of the practical hurdles, some physicists propose a redirection of high-energy physics away from accelerators to alternative empirical probes. For example, Hossenfelder (2019), in a *New York Times* opinion piece, argues against blindly constructing a new particle accelerator without having a clear idea of where to look. She advocates pursuing other avenues, including the astrophysical observations that provide evidence for the existence of dark matter.

Although Hossenfelder's piece was contentious (see for example the reply by Randall 2019), it is reflective of a broader trend in particle physics. It is generally the case that particle physicists have used the early universe more and more to look for constraints on BSM physics – albeit not necessarily as a replacement, but rather as complementary to accelerator experiments. This approach is essentially following through on Zel'dovich adage that the early universe is the 'poor man's accelerator'[2], or, in Rocky Kolb's words, the "ultimate particle accelerator" (Kolb 1998). The early universe reached energy levels far beyond anything achievable in terrestrial experiments. Thus, the hope is that observations

[1] Galison (1997) gives a comprehensive overview of experiments in particle physics in the twentieth century.

[2] Of course, the irony here is that constructing telescopes and experiments for observing the relics of the early universe is in no way cheap...

of relics of the early universe somehow contain signatures of BSM physics, or, at the very least, allow for the exclusion of some proposed BSM physics models.

This chapter takes a closer look at the use of the early universe as a new probe for BSM physics. The goal here is not to pronounce a recommendation on the debate about the future of particle physics. Rather, be it as a replacement for or as complementary to accelerator physics, the early universe probes to BSM physics represent a significant shift in the methodology of particle physics. Here, my focus is on what, if any, epistemic shift might be associated with that methodological shift. In other words, is the epistemic status of cosmic constraints on BSM physics somehow significantly different than that of constraints from accelerator experiments? In answering that question, I will achieve two things. First, I will show that existing philosophical accounts of the unique epistemic status of cosmological observations do not accurately pinpoint the epistemic shift associated with early-universe constraints. Second, I will begin carving out the proper locus of that epistemic shift: in a propagation of errors.

In the next section, I begin with the case of the early universe as the 'poor man's accelerator' (section 5.2). I specifically focus on Big Bang Nucleosynthesis (BBN), arguably the first constraint on BSM physics that was not produced in traditional laboratory experiments, and I explain how BBN signatures can be used to constrain BSM physics. That leads me to discussing two existing philosophical accounts on the unique status of empirical evidence in cosmology and astrophysics (section 5.3). I argue that these misidentify the locus of the epistemic shift for BSM constraints from cosmology. Instead, I use the neutron lifetime problem to introduce a new cautionary tale: a propagation of errors (section 5.4). In the conclusion, I briefly return to the debate on the future of particle physics.

Before I continue, note that the current focus of cosmological constraints on BSM physics has shifted partly to the power spectrum of the cosmic microwave background (CMB) anisotropies. This is in part because of limited precision in determining the primordial element abundances (cf. infra). This does not make the BBN case moot, however. First, the BBN case was historically the first source of constraints on BSM physics. Second, even though it is no longer the primary source of BSM constraints, research into BBN is ongoing. Finally, the BBN case provides a clear illustration of the philosophical argument, but the more general conclusions of this chapter extend beyond the BBN constraints.

5.2 THE EARLY UNIVERSE, OR A POOR MAN'S ACCELERATOR

Big Bang Nucleosynthesis (sometimes also referred to as primordial nucleosynthesis) is the theory of the formation of the lightest elements in the early universe. It was first developed in the 1940s by Alpher, Gamow and Herman (including in Alpher, Bethe, and Gamow 1948), and was independently adapted in the 1960s to modern cosmology by Peebles (1966) and Zel'dovich (1965). The theory draws on general relativity and nuclear physics to describe how hydrogen, deuterium, helium, and lithium were formed (see Weinberg 2008 for detailed theoretical derivations; see Cyburt et al. 2016; Particle Data Group 2018a for a summary of the theory and an overview of observational constraints). This section gives some general background on BBN and discusses how it has become one of the probes for particle physics beyond the standard model.

5.2.1 The basics of big bang nucleosynthesis

The cosmological context for BBN is an expanding universe with a Friedman-Lemaître-Robertson-Walker (FLRW) metric. The universe is assumed to be homogeneous and isotropic at the largest scales, and with curvature close to zero. During BBN, the universe was radiation-dominated: its expansion was primarily determined by the evolution of radiation (photons and neutrinos), rather than the evolution of matter or dark energy.[3]

BBN models basic nuclear physics in an expanding universe. Different stages of BBN are defined by the 'freeze out' of various nuclear reactions. Specifically, when a particular reaction rate falls below the expansion rate of the universe, the reaction effectively halts because the universe's expansion is pulling particles apart faster than that they can interact with one another.

[3]Different contributions to the energy density have different equations of state that describe the relation between their density and the scale factor of the universe. The universe is currently dark energy-dominated.

Prior to the onset of BBN (for a temperature $T > 10^{10}K$), proton/neutron-conversion reactions are in thermal equilibrium:

$$n + \nu_e \leftrightarrow p + e^- \tag{5.1}$$

$$p + \overline{\nu}_e \leftrightarrow n + e^+ \tag{5.2}$$

$$n \leftrightarrow p + e^- + \overline{\nu}_e \tag{5.3}$$

At this stage, the neutron-to-proton ratio is $X_n/X_p = exp\{\frac{-Q}{k_B T}\}$. In this equilibrium ratio, $Q = (m_n - m_p)c^2$.

At $10^{10}K > T > 3 * 10^9 K$, the two-body and three-body reactions freeze out. The neutron-to-proton ratio is approximately $X_n/X_p = 0.2$. Free neutron decay continues, however: since it does not require any two-body interactions, it can only be halted by neutrons no longer being free. Thus, the neutron-fraction continues to deteriorate: $X_n \sim exp\{\frac{-t}{\tau_n}\}$, with τ_n the neutron lifetime and t time. The conversion of neutrons to protons will only halt with the formation of complex nuclei containing neutrons.

Below $T = 10^9 K$, the formation of heavier nuclei commences. Due to the low density of the universe, nucleosynthesis can only occur through two-body reactions; higher-body reaction rates are too low to compete with the universe's expansion rate. The gateway to all heavier elements is thus the formation of deuterium:

$$n + p \rightarrow {}^2H + \gamma \tag{5.4}$$

This reaction fails to produce a net amount of deuterium until the inverse reaction, the photo-dissociation of deuterium, freezes out. Below $T = 0.7 * 10^9 K$, the number of photons per baryon with energy above the deuterium-dissociation threshold is less than 1 and the photo-dissociation reaction freezes out. The neutron-to-proton ratio thus decreases further to $X_n/X_p = 0.14$ due to ongoing free neutron decay.

After the deuterium-bottleneck is broken and the net production of deuterium begins, the other heavier elements are formed through a variety of two-body interactions:

$$^2H + n \rightarrow {}^3H + \gamma \tag{5.5}$$

$$^2H + p \rightarrow {}^3He + \gamma \tag{5.6}$$

$$^2H + {}^2H \rightarrow {}^3H + p \tag{5.7}$$

$$^2H + {}^2H \rightarrow {}^3He + n \tag{5.8}$$

$$^2H + {}^2H \rightarrow {}^4He + \gamma \tag{5.9}$$

$$^2H + {}^3H \rightarrow {}^4He + n \tag{5.10}$$

$$^2H + {}^3He \rightarrow {}^4He + p \tag{5.11}$$

$$^3H + p \rightarrow {}^4He + \gamma \tag{5.12}$$

$$^3He + n \rightarrow {}^4He + \gamma \tag{5.13}$$

(Although the one-deuterium-reactions commence before the photo-dissociation of deuterium freezes out, their cross-sections are intrinsically much smaller than two-deuterium processes.) Very few neutrons are lost to free neutron decay at this stage; they are almost all bound up in ^4He-nuclei. Some further reactions not listed here result in the formation of a small fraction of ^7Li. The production of heavier elements does not occur due to the lack of stable elements with mass number 5 or 8.

The empirical success of BBN is remarkable: its predictions for the relative primordial abundances of hydrogen, deuterium and helium match empirically determined abundances very closely. Empirically determining primordial abundances has to overcome an obvious problem: the primordial abundances were fixed billions of years before current observations. Between the end of BBN and the universe today, stellar nucleosynthesis has modified element abundances to a different degree in different regions of the universe. To determine primordial abundances, the focus is on regions with low metal-abundances, since any element heavier than ^7Li is produced in stellar nucleosynthesis. Abundances of the lighter elements in low-metal regions are assumed to be close-to-primordial. For example, the primordial helium abundance is partially determined based on recombination emission lines of helium and hydrogen in blue compact galaxies, whereas empirical determinations of the primordial lithium-abundance are based on metal-poor stars in the galactic spheroid. For all empirically determined abundances, uncertainties are currently dominated by systematic errors.

Aside from improving observational constraints, another concern for BBN is the lithium-problem. Given the empirically determined primordial abundances, measurements of nuclear parameters in terrestrial experiments, and fixed values for cosmological parameters based on CMB-observations, BBN allows for an internal consistency check. Specifically, all primordial abundances depend on the value of η, the primordial baryon-to-photon ratio. η determines when the photon-dissociation of deuterium freezes out, and therefore when the free neutron decay ends and the formation of heavier elements commences. Since most neutrons are bound up in ^4He-nuclei, more available neutrons implies a higher helium-abundance (commonly expressed in terms of the primordial helium mass fraction Y_p), and similarly for the other element abundances. The primordial ^4He- and ^2H-abundances generally align with the same value for η, which is also consistent with the value derived from CMB observations. The value for η derived from the ^7Li-abundance diverges from the others. How the lithium-problem can be resolved is an outstanding issue, and some suggest it might point to some new physics (cf. infra). Despite the lithium-problem, the general consensus is that BBN is remarkably successful.

5.2.2 Constraining new physics with BBN

Its overall success has made BBN a possible candidate to constrain particle physics beyond the standard model. The particle data group writes, for example:

> [BBN] is the most effective test for the cosmological viability of ideas concerning physics beyond the standard model. (Particle Data Group 2018a, p. 9)

The belief in BBN is not surprising. It is the earliest stage of cosmological expansion for which there is a broadly accepted and empirically supported model. Earlier stages, particularly inflation and baryogenesis, continue to be the subject of controversy, and neither is currently described by an empirically supported model.

The Particle Data Group (2018a) describes two ways in which BBN could point towards new physics. The first one is the aforementioned lithium-problem. Several classes of solutions for the lithium-problem exist. One is invoking a so-called ‘nuclear fix’, where otherwise subdominant nuclear reactions are somehow enhanced. Detailed lab measurements have largely excluded this possibility. Another option is an astrophysical solution, where some unknown stellar mechanism causes the destruction of lithium in the host halo stars. Although this option is still viable, it generally requires invoking *ad hoc* mechanisms and finetuning of stellar parameters, and it is therefore generally disfavoured. The final option is that the lithium problem reveals genuinely new physics at work during BBN. For example, the formation of light elements could be perturbed by the injection of hadronic or electromagnetic decay products of dark matter particles. Another proposal refrains from modifying the particle content of the early universe, but it changes the basic particle interactions by time varying the fundamental constants. Specifically, lowering the deuterium binding energy would lead the deuterium bottleneck to be broken later. This would lead to lower lithium-abundances but higher deuterium-abundances (since fewer deuterium nuclei would be bound up in heavier nuclei).

A second route from BBN to new physics uses observational bounds on the helium abundance, without introducing explicit modifications to the simple physics underlying BBN. BBN can be used to constrain the effective number of neutrino flavors N_{eff}. N_{eff} is a measure for the total number of relativistic particle species in the universe, that is, the total energy density in radiation in the universe excluding photons—it does not just cover the number of neutrino species. Changes in N_{eff} compared to the standard model value would affect the expansion rate of the universe. Since the expansion rate determines when different reactions freeze out, a change in expansion rate would affect the various abundances of the lightest elements. Based on observations from BBN, $N_{eff} = 3.28 \pm 0.28$ (Meyers 2018). Note that N_{eff} depends on η, which in this context is usually fixed by the deuterium abundance.

N_{eff}-constraints can be translated in specific constraints on the possibility of non-neutrino particle species that would somehow affect the expansion rate of the universe. Examples include sterile neutrinos, but also supersymmetric (SUSY) scenarios with a very light neutralino or gravitino. Finally, one of the main benefits of BBN is that it is complementary[4]

[4]‘Complementarity’ here has the same meaning as in context of dark matter particle searches discussed in chapter 2.0.

to accelerator data in constraining SUSY-parameter space.

Interestingly, N_{eff} can independently be determined based on the CMB anisotropy power spectrum. This determination similarly probes the radiation content of the universe, but at a much later stage in the universe's evolution. A discrepancy between the values from BBN and from the CMB could point towards the existence of so-called dark radiation (e.g. sterile neutrinos); the fact that the two values are currently consistent severely constrains this possibility.

5.3 PHILOSOPHERS ON EXPERIMENTS IN COSMOLOGY

In light of the recent debates about the future of high-energy physics (HEP), an epistemic analysis of the use of cosmological probes for BSM physics becomes pressing. Various philosophers of science have offered accounts of what makes cosmological and astrophysical observations in some way uniquely challenging compared to laboratory experiments. If their analyses hold, one might conclude that the methodological shift to the early universe as the ultimate particle accelerator also implies an epistemic shift in the status of empirical observations in HEP. Here, I examine two such previous accounts and I show that neither one properly extends to BSM constraints from the early universe.

5.3.1 Intervention, observation and bad jokes

Several recent views on the unique methodology of cosmology and/or astrophysics appeal to the distinction between observational and experimental sciences. Cosmology and astrophysics are supposed to qualify as purely observational sciences, where, due to the size and remoteness of the target, no active intervention in laboratory experiments is possible. It is then either considered evident or it is explicitly argued for that this difference between observation and intervention is somehow epistemically significant.

If this last claim is true, this would be relevant for the epistemic implications of the methodological shift to early universe probes for BSM physics. Past successes in high-energy physics have been a triumph of experimental science in controlled laboratory environments, with the discovery of the Higgs boson as its pinnacle. Observations of relics from the early universe decidedly do not qualify as experiments in a controlled laboratory environment. If there is a significant epistemic difference between intervention and observation, the epistemic status of early universe probes for BSM physics would therefore differ from that of accelerator results.

Thus, we are now faced with the question whether it is true that the difference between observation and intervention is somehow epistemically significant. In a recent presentation, Boyd (2019) reviews several accounts that classify cosmology and astrophysics as somehow distinctly observational, and shows that this distinction should not carry any weight with regards to the epistemology of cosmological and astrophysical practice. Here, I summarize Boyd's arguments and draw brief conclusions for the status of early universe constraints on BSM physics.

Although Boyd considers three different authors, Hacking (1989) is the only one that explicitly ties epistemic weight to the distinction between observation and intervention (building on Hacking 1983). He writes that "[g]alactic experimentation is science fiction, while extra-galactic experimentation is a bad joke" (Hacking 1989, p. 559). This lack of experimentation is supposedly devastating for the status of cosmology and astrophysics as empirical sciences: Hacking claims that the scientific method is characterized by experiments and interference with nature, or, in his words, "the creation of new phenomena" (ibid., p. 577), whereas astrophysics and cosmology merely save the phenomena. Hacking concludes that astrophysics and cosmology can therefore not properly qualify as scientific.

Boyd (2019) points out that Hacking's conclusions take an extreme normative position (perhaps reflective of a different era in cosmological and astrophysical practice). Taking Hacking's argument at face value would imply that the entire enterprise of modern cosmology and astrophysics is futile. There are several reasons to oppose this conclusion. First, Boyd (2019) discusses so-called "laboratory astrophysics", where laboratory experiments are designed to recreate the interior processes of supernovae. These types of experiments make a strong observation/intervention-distinction uneasy. Second, complementing Boyd's point about laboratory astrophysics, there are several examples in the history of HEP where astrophysical observations have led to advances in HEP theorizing. For example, observations of the solar neutrino flux and the subsequent solar neutrino problem led to the introduction of the idea of neutrino flavour oscillations (see Bahcall and Davis 1982 for a first-hand account). It is not obvious how these observations could qualify as 'proper' laboratory experiments, yet discarding them, as Hacking's account suggests, would mean discarding a swath of the recent history of HEP as somehow unscientific.

5.3.2 Overlap arguments and inferences from effects

A more recent paper by Weisberg et al. (2018) focuses on methodological challenges in the search for dark galaxies, galaxies completely composed of dark matter. Dark galaxies could strengthen the case for dark matter even further, but Weisberg et al. contend that it could also lead to a partial solution to the small-scale challenges to ΛCDM (see Bullock and Boylan-Kolchin 2017 for an overview of the small-scale challenges and possible solutions). Discovering such a dark galaxy faces a swath of challenges to establishing secure epistemic justification for a discovery. Ultimately, Weisberg et al. are optimistic about the possibility of finding sufficient epistemic justification for such a discovery, but they do believe that that kind of warrant might be particularly difficult to achieve. It is in that light that I look at their account: do the difficulties they identify raise worries about the status of cosmological probes for BSM physics?

Some of the difficulties that Weisberg et al. identify are specific to their particular project. For example, they rely heavily on citizen science and therefore question requirements of scientific expertise. A second challenge is that astronomy is an observational science—I have explained in the previous section why this does not obviously constitute an epistemic shift to the use of cosmological probes for BSM physics.

Most relevant here are two related challenges. First, there is the problem of finding an astrophysical object that is not detectable through the electromagnetic interaction.[5] Weisberg et al. claim that in astrophysics, the use of ever more powerful instruments is commonly

[5] Although I recognize this worry for finding a dark *galaxy*, I do believe that their description of dark matter not being detectable is slightly misleading. Although electromagnetic couplings have been excluded, it is still possible for dark matter particles to couple to standard model particles through other interactions. See previous chapters for more extensive discussions of this point.

justified using *overlap arguments*: arguments that show that a given phenomenon can be observed by a reliable experimental technique, as well as by a new technique of unknown reliability. The fact that the new technique overlaps with the old one then lends warrant to the reliability of the new technique. If that new technique detects an exotic new object, the discovery claim also receives partial epistemic warrant from the reliability of the instrument. The first challenge to observe dark galaxies is that there are no obvious overlap arguments available. The dark galaxy can't be detected through usual telescopic or spectroscopic instruments. To overcome this challenge, the presence of a dark galaxy will have to be inferred from its effects. But this leads to the second challenge: inferring the presence of new objects from their effects is quite common in scientific practice, but it is impeded in the dark galaxy-case due to serious under-theorizing (both with respect to dark matter as well as the potential effects of dark galaxies). This means that it is not obvious what the observational effects could be of the presence of a dark galaxy, and what observations would therefore suffice to make a discovery claim about such a dark galaxy.

At first sight, the two challenges faced by the dark galaxy hypothesis seem to apply to the cosmic constraints on BSM physics as well. Since it is not clear what the coupling of BSM physics to SM particles might be, it is not obvious what instruments would be useful to detect it. Therefore, the presence of BSM physics in the early universe, like that of a dark galaxy, can only be inferred from its effects, for example on the primordial abundances of the lightest elements or on the CMB anisotropies. But BSM physics is under-theorized in the same way that the dark galaxy hypothesis is under-theorized: although many theoretical proposals for BSM physics exist, no single one can clearly be preferred.

Despite superficial similarities to the dark galaxy case, the problem of inferring the presence of an undertheorized exotic new object from its effects does not pinpoint the epistemic shift associated with early universe constraints on BSM physics. A main difference between the dark galaxy search and the BSM search is the nature of the potential discovery. Weisberg et al.'s goal is to establish a specific and particular discovery, namely the discovery of a particular dark galaxy in a specific location in the universe. This is much more ambitious than the goal of the cosmological probes for BSM physics: currently, they are merely trying to constrain the space of possibilities for BSM physics. That happens through a process of elimination, in other words, looking for any kind of effect that cannot be explained by existing standard model physics. Because the current goals of early universe probes on BSM physics are more minimal than those of the dark galaxy project, the lack of overlap arguments is not yet a concern. The current observations all still use instruments for which overlap arguments are available.

With regards to the second challenge, the fact that BSM physics-effects in the early universe are undertheorized is also not significantly different compared to terrestrial experiments. In fact, McCoy and Massimi (2018) extensively discuss the use of exploratory models in BSM-searches at the LHC. Whether or not BSM physics is undertheorized will thus affect early universe-constraints and accelerator-experiments equally.

5.4 OUT OF THE FYRING PAN, INTO THE FIRE?

So far, I have discussed the overall success of BBN as a theory of the formation of primordial elements, and I have argued that existing philosophical concerns about the epistemic status of observations in cosmology and astrophysics don't extend to the use of early universe probes for constraining BSM physics. Does that mean that, despite the methodological shift, there is no obvious epistemic shift associated with the early universe as a replacement for terrestrial accelerators? Unsurprisingly, the answer I offer here is: 'not quite'. A recent controversy about the neutron lifetime reveals the proper locus of concern for early universe-probes of BSM physics. I first discuss the specifics of the neutron lifetime case, and I then draw conclusions for the epistemic shift associated with early universe-probes of BSM physics.

5.4.1 The neutron lifetime

The neutron lifetime τ_n plays a crucial role in BBN. Kolb and Turner write that "[a] high precision measurement of τ_n [...] will all but eliminate the uncertainty in the predicted ^4He primordial abundance" (Kolb and Turner 1988, p. 110). Almost three decades later, the Particle Data Group similarly states that "[t]he experimental parameter most important in determining Y_p [the helium-to-hydrogen ratio] is the neutron lifetime, τ_n" (Particle Data Group 2018a, p. 2).

There are two ways in which the neutron lifetime plays a crucial role in BBN. First, it determines the reaction rates for neutron-proton conversion reactions, and it therefore determines when these reactions freeze out and the formation of deuterium begins. Second, it determines the reaction rate for free neutron decay, which only halts once the deuterium bottle-neck has been broken. τ_n thus determines the neutron-to-proton ratio in the early universe, and, since almost all available free neutrons in the early universe are captured in ^4He-nuclei, the primordial helium abundance Y_p.

Of course, τ_n is not the only free parameter determining freeze-out of various reactions in the early universe; above, I also discussed the importance of the effective number of neutrino species N_{eff} and the baryon-to-photon ratio η. In an ideal scenario, all three parameter values can be fixed based on alternative observations, such that precise BBN predictions can be compared to observations. In the past, various authors have proclaimed this to be possible, and they have therefore declared BBN "a parameter-free theory" (Dubbers and Schmidt 2011, p. 72; see also Wietfeldt and Greene 2011; Dubbers 2005).

This declaration only holds insofar as the experimentally determined values for the three parameters hold. This is where the problem arises: recent experimental determinations of τ_n have started to disagree. As discussed in chapter 2.0, there are two main methods to measure τ_n: trapping neutrons in a gravito-magnetic 'bottle' (Pattie et al. 2018) or counting decay products along a neutron beam (Yue et al. 2013). The two measurement protocols predict different values for τ_n. Where beam experiments measure $\tau_n = 886,3 \pm 1.2(stat) \pm 3.2(syst)\, s$, bottle experiments measure $\tau_n = 877,7 \pm 0.7(stat) + 0.4 - 0.2(syst)\, s$. This controversy started over 15 years ago, with no clear resolution in sight (see Witze 2019 for a report on a recent workshop on the issue).

Regardless of the source of the discrepancy, it does raise an important question for BBN: what value for τ_n should be used in calculating primordial element abundances? One option is to follow the Particle Data Group. In their BBN calculations, they use a weighted average of seven recent measurements of τ_n: $\tau_n = 880,2 \pm 1.0\, s$ (Particle Data Group 2018b, pp. 2-3).

The seven values come from double bottle experiments (1), material bottle experiments (3), gravitational trap experiments (1), Penning trap experiments (1) and beam measurements (1).[6] This pragmatic solution sits uneasy, however: while the Particle Data Group report does recommend a single average value, their weighted average suggests a much broader probability distribution (Particle Data Group 2018b, p. 4).

Given the crucial role of τ_n in BBN, the controversy surrounding τ_n can not obviously be cast aside as unimportant for BBN. The importance of τ_n is further reinforced by a recent update to AlterBBN, an open source code from 2011 for calculating primordial element abundances (see Arbey 2012 for the original manual, see Arbey et al. 2019 for the update). The update introduced several innovations, *including* the possibility to alter the value for τ_n and compare the resulting predicted abundances with empirically determined values.

Let me not overstate the concern. First, as mentioned before, there are alternative constraints on BSM physics available from the CMB (see Planck Collaboration 2018 for details). The CMB power spectrum can constrain N_{eff}, as well as the neutrino masses independently from BBN. Second, an alternative approach uses CMB-derived values for standard cosmological parameters (ΛCDM has only six free parameters, all of which can be derived from the CMB anisotropy power spectrum) and plugs those into the standard BBN scenario in order to make detailed predictions for primordial element abundances as a consistency check for ΛCDM. The predicted abundances are consistent with various determinations of Y_p (including two disagreeing values), and the uncertainties on τ_n are small enough to not significantly affect the derivations. Finally, given the CMB constraints and the standard scenario for BBN, BBN can be used to provide an independent measure for τ_n that might help resolve the current tension between bottle- and beam-experiments. Here, however, the discrepancies between the different observational determinations of Y_p are possible cause for concern.

5.4.2 The new cautionary tale, completed

The neutron lifetime discrepancy might be resolved in the next couple of years through improved experiments or error estimations. Similarly, observational constraints on primordial

[6]This average does not include the most recent bottle measurements from Pattie et al. (2018), which would lower the average lifetime even further.

element abundances may see significant improvement as more and more systematics are accounted for. Regardless of these future developments, the case of the neutron lifetime discrepancy and its role in BBN indicates a possible epistemic shift associated with the methodological shift from accelerator to early universe-probes for BSM physics.

Because early-universe constraints on BSM physics are embedded in the complete story about the evolution of the universe from a hot dense state to the universe today, the physical processes that are supposedly relevant to account for the observations extend over immense spatio-temporal scales. On the one hand, the processes where new physics might actively play a role (BBN, the universe's expansion rate in its radiation-dominated era) extend on the scale of the entire universe, in combination with gravity and standard model-physics. On the other hand, the spatio-temporal scales relevant for the observations of relics of these aforementioned processes (primordial element abundances, CMB) are also significantly larger than those for terrestrial experiments.

As a result, early universe-probes for BSM physics present both an opportunity, and a threat. The opportunity is that there can be several independent sources of evidence for BSM physics, since BSM physics can have had multiple effects in the early universe. Indeed, both primordial element abundances and the CMB might reveal effects of BSM physics (to a variable degree of independence). This is all the more important if BSM physics is hiding at scales inaccessible to terrestrial accelerators.

The threat is that, because so many more physical processes become relevant for the production of signals of BSM physics and for the conservation of those relics today, there is a danger of an accumulation of errors. For example, in determining the presence of BSM physics from primordial element abundances, there are concerns about the modeling of systematic errors in those observations, as is the case for estimates of Y_p. But in addition to those usual concerns about systematics in observations, there are additional concerns about uncertainties in presumed 'established' quantities like the neutron lifetime. This second source of uncertainties comes out of the theoretical assumptions that are required for modelling the effects of standard model physics. In other words, errors from disciplines that were initially informing the cosmological model now come back to bite when the cosmological evidence is interpreted in ways that constrain those disciplines.

Is it obvious that this threat is a difference in kind from concerns about systematics errors influencing the modelling of background physics in accelerator experiments? Perhaps not in principle. After all, accelerator experiments require an immense amount of modelling and data processing in order to control for systematic errors in the detectors and to model the background against which new physics could be detected. In the current status of the field, there seems to be a significant practical difference between accelerator experiments and early universe constraints, however, just because of the sheer scale at which early universe constraints are produced and observed.

5.5 CONCLUSION

This chapter started with an overview of possible future experiments in HEP, and the debate as to whether a new and improved accelerator should be built, or forsaken in favor of early universe probes for BSM physics. The main goal here was to investigate whether there is a significant epistemic shift associated with the proposed methodological shift for HEP. I argued that existing philosophical views on what makes cosmological observations different from terrestrial experiments do not accurately locate the epistemic shift. Rather, if there is such an epistemic shift, it lies primarily in the taking on of existing systematic errors in the required modelling of SM physics to find evidence of BSM physics, more than anything else.

6.0 Concluding Remarks: Lessons for Integration in Philosophy of Science

Cosmology sits at the intersection of various disciplines, one of them particle physics. This dissertation has specifically focused on the exchange between particle physics and cosmology, both in context of dark matter research and research on physics beyond the standard model (areas that are, of course, not mutually exclusive). After exploring various epistemological issues, let me return to the general question of integration in science. My main focus has been on experimental practice, and from that focus, I am now in a position to draw some lessons about integrative science.

Lesson 1—Integration as an alternative perspective on inter-theory relations. A first lesson is that cosmology provides another example of how integration is an alternative to other approaches to inter-theory relations. First, the fact that integration responds to just one (albeit complex) problem sets it apart from multidisciplinarity and transdisciplinarity, which are usually mentioned in the same context as integration (see for example Klein 1990; Hoffmann, Schmidt, and Nersessian 2013; Grüne-Yanoff 2016). In multidisciplinary research, multiple disciplines work simultaneously towards a common goal, but the disciplines do not cooperate to provide one answer. They remain juxtaposed, rather than contributing to an integrative model. Integration is also different from transdisciplinarity, the process where there is genuine integration of disciplines, but where that integration has a goal that lies outside of science.

Second, the 'coupling back' in integration should not be understood as integration moving towards unification. ΛCDM is no unification of general relativity and particle physics under a limited set of brute facts; rather, it is a model of the universe that draws on several disparate theoretical frameworks to model different aspects of the evolution of the universe. It is true that ΛCDM does 'unify' a variety of cosmological phenomena under a small set of equations with six free parameters. This is different, however, from unification at the level of the standard model of particle physics and general relativity: the different theoretical frameworks maintain their independence from one another, despite the cosmological integration. In other words, I follow Mitchell (2003)'s integrative pluralism in that pluralism at the theoretical level is warranted (although I have not explicitly argued for this in the dissertation), and that integration is required only at the concrete level, when attempting to model the evolution of the universe.

Third, and related, integration should also be distinguished from any notion of reduction. Again, ΛCDM is not redescribing cosmological phenomena in terms of a single theory of fundamental physics. Rather, it is built up out of non-hierarchical interactions between disciplines, many of which are considered 'fundamental physics' independently of one another, like particle physics and general relativity. Mitchell (2003) and Batterman (2016) provide further arguments against including all types of inter-theory relations under just two umbrellas of unification and reduction. The intersection between particle physics and cosmology provides an example from practice in support of this idea.

Lesson 2—The empirical aspects of integration. A second lesson is that integration does not necessarily have to be focused on integrative model-building. An important aspect of integrative science is empirical. In chapter 2.0, I introduced the method-driven logic. The case of the particle dark matter searches showed that, even in cases where the model is underdeveloped and explanatory integration has not yet been achieved, there can be a type of methodological integration. Similarly, the conflicts in evidence from chapter 4.0 and the 'coupling back' discussed in chapter 5.0 are specific instances of the integration of different lines of evidence.

There is a risk associated with framing integration purely in terms of theoretical model-building, in that this might eclipse other aspects of integrative practice—those aspects that are associated with experiments and empirical data. I follow Potochnik (2011) and Plutynski (2013) that the notion of 'integration' in philosophy of science should allow for broader discussion of non-derivative connections between disciplines. Thus, it should also encompass experimental integrative practices. One example is Currie (2018)'s account of 'methodological omnivory' in paleontology; this dissertation is another.

Lesson 3—Integration is a two-way street. A third lesson is that integration is not a one-way street, something that has not been discussed extensively in the existing philosophical literature. In chapter 5.0, I described epistemic issues associated with a 'coupling back' from cosmology to some of its informative disciplines. There, the coupling back takes the form of providing new lines of evidence for particle physics. Schneider (2020)'s discussion of the cosmological constant problem might provide an example of such coupling back on the theoretical level. More broadly, Ellis (2018) mentions the inclusion of cosmology in the reports of the Particle Data Group as evidence that cosmology today contributes to a large extent to other fields of physics. Coupling back is often met with a unique set of challenges, and those challenges are importantly different form those faced by the initial integrative effort. This coupling back thus merits further exploration.

My goal here has not been to offer a full-fledged account of integrative science, perhaps in part because integration seems to be a catch-all for relations between disciplines that are not reductive or unificationist. But that does not mean that there are not some more specific lessons to be drawn for examples of integrative science. Here, I have offered some, with regards to how such non-reductive, non-unificatory connections can arise in experimental practice, and how they can also result in a coupling-back.

7.0 Bibliography

Albrecht, Andreas et al. (2006). “Report of the Dark Energy Task Force”. In: arXiv: 0609591 [astro-ph].

Alpher, R A, H Bethe, and G Gamow (1948). “The Origin of Chemical Elements”. In: *Physical Review* 73.7, pp. 803–04.

Andersen, Hanne and Brian Hepburn (2016). “Scientific Method”. In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Summer 2016. Metaphysics Research Lab, Stanford University. URL: https://plato.stanford.edu/archives/sum2016/entries/scientific-method/.

Arbey, A et al. (2019). “AlterBBN v2: A Public Code for Calculating Big-Bang Nucleosynthesis Constraints in Alternative Cosmologies”. In: arXiv: 1806.11095v2.

Arbey, Alexandre (2012). “AlterBBN: A Program for Calculating the BBN Abundances of the Elements in Alternative Cosmologies”. In: *Computer Physics Communications* 183.8, pp. 1822–31.

Baade, Walter (1956). “The Period-Luminosity Relation of the Cepheids”. In: *The Astronomical Journal* 68, pp. 5–16.

Bahcall, John N and Raymond Davis (1982). “An Account of the Development of the Solar Neutrino Problem”. In: *Essays in Nuclear Astrophysics*. Ed. by Charles A. Barnes, Donald D. Clayton, and David Schramm. Cambridge: Cambridge University Press, pp. 243–85.

Bahcall, Neta A. (1995). “Rating Cosmological Models”. In: arXiv: 9503110 [astro-ph].

— (1996). “Large-Scale Structure in the Universe”. In: *New Horizons from Multi-Wavelength Sky Surveys, Proceedings of the 179th Symposium of the International Astronomical Union, held in Baltimore, USA August 26-30, 1996*. Ed. by Brian J. McLean et al. Kluwer Academic Publishers, p. 317.

Banks, Michael (2019). *Disappointment as Japan Fails to Commit to Hosting the International Linear Collider*. URL: https://physicsworld.com/a/disappointment-as-japan-fails-to-commit-to-hosting-the-international-linear-collider/ (visited on 03/07/2019).

Barrow, John (1993). "Anthropic Principles in Cosmology". In: *Vistas in Astronomy* 37.C, pp. 409–27.

Batterman, Robert (2016). "Intertheory Relations in Physics". In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edouard Zalta. Fall 2016. Metaphysics Research Lab, Stanford University. URL: https://plato.stanford.edu/archives/fall2016/entries/physics-interrelate/.

Beisbart, Claus (2009). "Can We Justifiably Assume the Cosmological Principle in Order to Break Model Underdetermination in Cosmology?" In: *Journal for General Philosophy of Science* 40.2, pp. 175–205.

Beisbart, Claus and Tobias Jung (2006). "Privileged, Typical, or Not Even That? Our Place in the World According to the Copernican and the Cosmological Principles". In: *Journal for General Philosophy of Science* 37.2, pp. 225–56.

Belgiorno, F et al. (2010). "Hawking Radiation from Ultrashort Laser Pulse Filaments". In: *Physical Review Letters* 105, p. 203901.

Benétreau-Dupin, Yann (2015). "The Bayesian Who Knew Too Much". In: *Synthese* 192, pp. 1527–42.

Bernal, José and John A Peacock (2018). "Conservative Cosmology: Combining Data with Allowance for Unknown Systematics". In: *Journal of Cosmology and Astroparticle Physics,* 07, p. 002.

Bertone, Gianfranco and Dan Hooper (2016). "History of Dark Matter". In: *Reviews of Modern Physics* 90, p. 045002.

Bertone, Gianfranco, Dan Hooper, and Joseph Silk (2005). "Particle Dark Matter: Evidence, Candidates and Constraints". In: *Physics Reports* 405, pp. 279–390.

Bogen, James (2017). "Theory and Observation in Science". In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Summer 2017. Metaphysics Research Lab, Stanford University. URL: https://plato.stanford.edu/archives/sum2017/entries/science-theory-observation/.

Boran, S et al. (2018). “GW170817 Falsifies Dark Matter Emulators”. In: *Physical Review D* 97.4, p. 41501.

Bouchendira, Rym, Francois Nez, and Francois Biraben (2013). “State of the Art in the Determination of the Fine Structure Constant: Test of Quantum Electrodynamics and Determination of h/mu”. In: *Annalen der Physik* 525.7, pp. 484–92.

Boyd, Nora Mills (2018). “Evidence Enriched”. In: *Philosophy of Science* 85.July, pp. 403–21.

— (2019). *Astrophysical Evidence: Observation and Intervention are Irrelevant*. Universidad Complutense Madrid.

Brandenberger, Robert (2014). “Do We Have a Theory of Early Universe Cosmology?” In: *Studies in History and Philosophy of Science Part B: History and Philosophy of Modern Physics* 46.1, pp. 109–21.

British Association for the Advancement of Science (1932). *Report of the British Association for the Advancement of Science*. August. London, p. 736.

Buchmueller, Oliver, Caterina Doglioni, and Lian-Tao Wang (2017). “Search for Dark Matter at Colliders”. In: *Nature Physics* 13.3, pp. 217–23.

Bullock, James S. and Michael Boylan-Kolchin (2017). “Small Scale Challenges to the ΛCDM Paradigm”. In: *Annual Review of Astronomy and Astrophysics* 55, pp. 343–87.

Burrage, Clare, Edmund Copeland, and E A Hinds (2015). “Probing Dark Energy with Atom Interferometry”. In: *Journal of Cosmology and Astroparticle Physics* 3, p. 042.

Butterfield, Jeremy (2014). “On Under-Determination in Cosmology”. In: *Studies in History and Philosophy of Science Part B: History and Philosophy of Modern Physics* 46.1, pp. 57–69.

Castelvecchi, Davide (2018). “Dark-matter Detector in Italy Strikes Again”. In: *Nature* 556, pp. 13–14.

Cirkovic, Milan M (2002). “Anthropic Fluctuations vs. Weak Anthropic Principle”. In: *Foundations of Science* 1917, pp. 453–63.

Ćirković, Milan M. and Slobodan Perović (2016). “Alternative Explanations of the Cosmic Microwave Background: A Historical and an Epistemological Perspective”. In: *Studies in*

History and Philosophy of Science Part B: History and Philosophy of Modern Physics 62, pp. 1–18.

CLIC and CLICdp (2018). “The Compact Linear e+ e- Collider (CLIC): Physics Potential”. In: arXiv: `1812.07986v1`.

Clowe, Douglas et al. (2006). “A Direct Empirical Proof of the Existence of Dark Matter”. In: *The Astrophysical Journal* 648, pp. L109–13.

Colaço, David (2018). “Rethinking the Role of Theory in Exploratory Experimentation”. In: *Biology & Philosophy* 33.5, pp. 1–17.

Conrad, Jan and Olaf Reimer (2017). “Indirect Dark Matter Searches in Gamma and Cosmic Rays”. In: *Nature Physics* 13.3, pp. 224–31.

Crowther, Karen, Niels S Linnemann, and Christian Wüthrich (2019). “What We Cannot Learn from Analogue Experiments”. In: *Synthese* Forthcoming.

Currie, Adrian M. (2018). *Rock, Bone, and Ruin. An Optimist’s Guide to the Historical Sciences*. Cambridge: The MIT Press, p. 376.

Cyburt, Richard H et al. (2016). “Big Bang Nucleosynthesis: Present Status”. In: *Reviews of Modern Physics* 88.1, p. 015004.

Dandapani, B. and M. Fleischmann (1972). “Electrolytic Separation Factors on Platinum Under Non-Steady State Conditions”. In: *Electroanalytical Chemistry and Interfacial Electrochemistry* 39, pp. 315–21.

Dardashti, Radin, Karim Thébault, and Eric Winsberg (2017). “Confirmation via Analogue Simulation: What Dumb Holes Could Tell Us about Gravity”. In: *British Journal for the Philosophy of Science* 68, pp. 55–89.

Darden, Lindley and Nancy Maull (1977). “Interfield Theories”. In: *Philosophy of Science* 44.1, pp. 43–64.

Dawid, Richard (2013). *String Theory and the Scientific Method*. Cambridge: Cambridge University Press, p. 202.

De Swart, J. G., G. Bertone, and J. Van Dongen (2017). “How Dark Matter Came to Matter”. In: *Nature Astronomy* 1, p. 0059.

Dubbers, Dirk (2005). “Particle Physics with Cold Neutrons”. In: *AIP Conference Proceedings* 769.9, pp. 9–13.

Dubbers, Dirk and Michael G Schmidt (2011). “The Neutron and its Role in Cosmology and Particle Physics”. In: *Reviews of Modern Physics* 83.4, pp. 1111–71.

Earman, John (1995). *Bangs, Crunches, Whimpers, and Shrieks: Singularities and Acausalities in Relativistic Spacetime*. Oxford: Oxford University Press, p. 257.

Earman, John and Jesus Mosterintl (1999). “A Critical Look at Inflationary Cosmology”. In: *Philosophy of Science* 66.1, pp. 1–49.

Eddington, Arthur (1930). “On the Instability of Einstein’s Spherical World”. In: *Monthly Notices of the Royal Astronomical Society* 49.May, pp. 668–78.

Einstein, Albert (1917). “Kosmologische Betrachtungen zur allgemeinen Relativitätstheorie”. In: *Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften* Februari, pp. 142–52.

Einstein, Albert and Willem de Sitter (1932). “On the Relation Between the Expansion and the Mean Density of the Universe”. In: *Proceedings of the National Academy of Sciences* 18.3, pp. 213–14.

Ellis, George (2018). “The Standard Cosmological Model: Achievements and Issues”. In: *Foundations of Physics* 48.10, pp. 1226–45.

Fleischmann, Martin and Stanley Pons (1989). “Electrochemically Induced Nuclear Fusion”. In: *Journal for Electronalytic Chemistry* 261, pp. 301–08.

Franklin, Laura R. (2005). “Exploratory Experimentation”. In: *Philosophy of Science* 72.5, pp. 888–99.

Freese, Katherine, Mariangela Lisanti, and Christopher Savage (2012). “Annual Modulation of Dark Matter: A Review”. In: *Reviews of Modern Physics* 85, p. 1561.

Galison, Peter (1997). *Image and Logic: A Material Culture of Microphysics*. Chicago: The University of Chicago Press, p. 982.

Gelmini, Graciela B., Lawrence J. Hall, and M. J. Lin (1987). “What is the Cosmion?” In: *Nuclear Physics B* 281, pp. 726–35.

Goodman, Mark W. and Edward Witten (1985). “Detectability of Certain Dark-Matter Candidates”. In: *Physical Review D* 31.12, pp. 3059–63.

Grüne-Yanoff, Till (2016). “Interdisciplinary Success Without Integration”. In: *European Journal for Philosophy of Science* 6.3, pp. 343–60.

Guala, Francesco (2003). “Experimental Localism and External Validity”. In: *Philosophy of Science* 70.December, pp. 1195–1205.

Gueguen, Marie (2020). “On Robustness in Cosmological Simulations”. In: *Philosophy of Science* Forthcoming.

Hacking, Ian (1983). *Representing and Intervening*. Cambridge: Cambridge University Press, p. 287.

— (1989). “Extragalactic Reality: The Case of Gravitational Lensing”. In: *Philosophy of Science* 56.4, pp. 555–81.

Hamilton, Paul et al. (2015). “Atom-Interferometry Constraints on Dark Energy”. In: *Science* 349.6250, pp. 849–51.

Hoffmann, Michael H G, Jan C. Schmidt, and Nancy J. Nersessian (2013). “Philosophy Of and As Interdisciplinarity”. In: *Synthese* 190.11, pp. 1857–64.

Hong, Tae Min (2017). “Dark Matter Searches at the LHC”. In: arXiv: `1709.02304`.

Hossenfelder, Sabine (2019). *The Uncertain Future of Particle Physics*. URL: `https://www.nytimes.com/2019/01/23/opinion/particle-physics-large-hadron-collider.html` (visited on 01/23/2019).

Hubble, Edwin (1929). “A Relation Between Distance and Radial Velocity Among Extra-Galactic Nebulae”. In: *Proceedings of the National Academy of Sciences* 15.3, pp. 168–73.

Jaffe, Matt et al. (2017). “Testing Sub-Gravitational Forces on Atoms from a Miniature In-Vacuum Source Mass”. In: *Nature Physics* 13.October, pp. 938–42.

Karaca, Koray (2017). “A Case Study in Experimental Exploration: Exploratory Data Selection at the Large Hadron Collider”. In: *Synthese* 194.2, pp. 333–54.

Khoury, Justin et al. (2001). “The Ekpyrotic Universe: Colliding Branes and the Origin of the Hot Big Bang”. In: *Physical Review D* 64, p. 123522.

Kincaid, Harold (1990). “Molecular Biology and the Unity of Science”. In: *Philosophy of Science* 57.4, pp. 575–93.

— (1997). *Individualism and the Unity of Science*. Oxford: Rowman & Littlefield Publishers.

Klein, Julie Thompson (1990). *Interdisciplinarity: History, Theory, and Practice*. Detroit: Wayne State University Press, p. 331.

Kolb, Edward W (1998). *Particle Physics in the Early Universe*. URL: https://ned.ipac.caltech.edu/level5/Kolb/paper.pdf (visited on 08/28/2019).

Kolb, Edward W. and Michael S. Turner (1988). *The Early Universe*. Redwood City: Addison Wesley, p. 547.

Kragh, Helge (2009). "Contemporary History of Cosmology and the Controversy over the Multiverse". In: *Annals of Science* 66.4, pp. 529–51.

— (2013). "Testability and Epistemic Shifts in Modern Cosmology". In: *Studies in History and Philosophy of Science Part B: History and Philosophy of Modern Physics* 46, pp. 48–56.

Laudan, Larry and Jarrett Leplin (1991). "Empirical Equivalence and Underdetermination". In: *The Journal of Philosophy* 88.9, pp. 449–72.

Lemaître, Georges (1927). "Un Univers Homogène de Masse Constante et de Rayon Croissant, Rendant Compte de la Vitesse Radiale des Nébuleises Extra-Galactiques". In: *Annales de la Société Scientifiques de Bruxelles, serie A: Sciences Mathématiques* XLVII.1, pp. 49–59.

— (1931). "The Expanding Universe". In: *Monthly Notices of the Royal Astronomical Society* 91.5, pp. 490–501.

— (1949). "The Cosmological Constant". In: *Albert Einstein: Philosopher Scientist. The Library of Living Philosophers VII*. Ed. by Paul Arthur Schilpp. Evanston, pp. 439–56.

Leonelli, Sabina (2013). "Integrating Data to Acquire New Knowledge: Three Modes of Integration in Plant Science". In: *Studies in History and Philosophy of Biological & Biomedical Sciences* 44.4, pp. 503–14.

— (2016). *Data-Centric Biology. A Philosophical Study*. Chicago: The University of Chicago Press, p. 275.

Liu, Jianglai, Xun Chen, and Xiangdong Ji (2017). "Current Status of Direct Dark Matter Detection Experiments". In: *Nature Physics* 13.3, pp. 212–16.

Love, Alan (2008). "Explaining Evolutionary Innovations and Novelties: Criteria of Explanatory Adequacy and Epistemological Prerequisites". In: *Philosophy of Science* 75.5, pp. 874–86.

Love, Alan and Gary Lugar (2013). "Dimensions of Integration in Interdisciplinary Explanations of the Origin of Evolutionary Novelty". In: *Studies in History and Philosophy of Science Part C: History and Philosophy of Biological and Biomedical Sciences* 44.4, pp. 537–50.

Massimi, Michela (2018). "Perspectival Modeling". In: *Philosophy of Science* 85.July, pp. 335–59.

— (2019). "Two Kinds of Exploratory Models". In: *Philosophy of Science* 86.5, pp. 869–81.

McCoy, Casey (2015). "Does Inflation Solve the Hot Big Bang Model's Fine-Tuning Problems?" In: *Studies in History and Philosophy of Science Part B: History and Philosophy of Modern Physics* 51, pp. 23–36.

— (2017). "Can Typicality Arguments Dissolve Cosmology's Flatness Problem? The Gibbons-Hawking-Stewart Measure". In: *Philosophy of Science* 84.5, pp. 1239–52.

McCoy, Casey and Michela Massimi (2018). "Simplified Models: A Different Perspective on Models as Mediators". In: *European Journal for Philosophy of Science* 8, pp. 99–123.

McGaugh, Stacy (2015). "A Tale of Two Paradigms: The Mutual Incommensurability of LCDM and MOND". In: *Canadian Journal of Physics* 93, pp. 250–59.

McMullin, Ernan (1993). "Indifference Principle and Anthropic Principle in Cosmology". In: *Studies in History and Philosophy of Science Part A* 24.3, pp. 359–89.

Merrit, David (2017). "Cosmology and Convention". In: *Studies in History and Philosophy of Science Part B: History and Philosophy of Modern Physics* 57, pp. 41–52.

Meyers, Joel (2018). *Cosmological Probes of Fundamental Physics*. URL: `http://www.physics.smu.edu/web/seminars/seminar{\_}archive/Spring2018/Joel{\_}Meyers.pdf` (visited on 08/28/2019).

Milgrom, M (1983). "A Modification of the Newtonian Dynamics as a Possible Alternative to the Hidden Mass Hypothesis". In: *Astrophysical Journal* 270, pp. 365–70.

Mitchell, Sandra (2003). *Biological Complexity and Integrative Pluralism*. Cambridge: Cambridge University Press, p. 244.

Mitchell, Sandra D and Angela M Gronenborn (2017). "After Fifty Years, Why Are Protein X-ray Crystallographers Still in Business?" In: *British Journal for the Philosophy of Science* 68, pp. 703–23.

Narlikar, Jayant V (2018). "The Evolution of Modern Cosmology as Seen Through a Personal Walk Across Six Decades". In: *The European Physical Journal H* 43.1, pp. 43–72.

Norton, John (2018). "Replicability of Experiment". In: *The Material Theory of Induction.* Pittsburgh. URL: `https://www.pitt.edu/{~}jdnorton/papers/Replicability{\_}experiment{\_}final.pdf`.

O'Malley, Maureen A. (2013). "When Integration Fails: Prokaryote Phylogeny and the Tree of Life". In: *Studies in History and Philosophy of Science Part C: History and Philosophy of Biological and Biomedical Sciences* 44.4, pp. 551–62.

O'Malley, Maureen A. and Orkun S. Soyer (2012). "The Roles of Integration in Molecular Systems Biology". In: *Studies in History and Philosophy of Science Part C: History and Philosophy of Biological and Biomedical Sciences* 43.1, pp. 58–68.

Parke, Emily C. (2014). "Experiments, Simulations and Epistemic Privilege". In: *Philosophy of Science* 81.October, pp. 516–36.

Parker, Richard et al. (2018). "Measurement of the Fine-Structure Constant as a Test of the Standard Model". In: *Science* 360, pp. 191–95.

Parker, Wendy S. (2009). "Does Matter Really Matter? Computer Simulations, Experiments, and Materiality". In: *Synthese* 169.3, pp. 483–96.

Particle Data Group (2018a). "Big-Bang Nucleosynthesis". In: *Physical Review D* 98.030001.

— (2018b). "n****". In: *Physical Review D* 98.030001.

— (2019a). "Dark Energy". In: *Physical Review D* 98.030001.

— (2019b). "Dark Matter". In: *Physical Review D* 98.030001.

Pattie, R W et al. (2018). "Measurement of the Neutron Lifetime Using a Magneto-Gravitational Trap and In Situ Detection". In: *Science* 360, pp. 627–32.

Peebles, P. James E (1966). "Primeval Helium Abundance and the Primeval Fireball". In: *Physical Review Letters* 16.10, pp. 410–13.

Peebles, P. James E., Lyman A. Page Jr., and R. Bruce Partridge, eds. (2009). *Finding the Big Bang.* Cambridge: Cambridge University Press, p. 571.

Penrose, Roger (2014). "On the Gravitization of Quantum Mechanics 2: Conformal Cyclic Cosmology". In: *Foundations of Physics* 44, pp. 873–90.

Penzias, Arno A. and Robert W. Wilson (1965). "A Measurement of Excess Antenna Temperature at 4080 Mc/s". In: *The Astrophysical Journal* 142.1, pp. 419–21.

Perlmutter, S et al. (1999). "Measurements of Omega and Lambda from 42 High-Redshift Supernovae". In: *The Astronomical Journal* 517, pp. 565–86.

Planck Collaboration (2018). "Planck 2018 results. VI. Cosmological Parameters". In: *Astronomy & Astrophysics*, pp. 1–71.

Plutynski, Anya (2013). "Cancer and the Goals of Integration". In: *Studies in History and Philosophy of Science C: History and Philosophy of Biological & Biomedical Sciences* 44.4, pp. 466–76.

Potochnik, Angela (2011). "A Neurathian Conception of the Unity of Science". In: *Erkenntnis* 74.3, pp. 305–19.

Primack, Joel R., David Seckel, and Bernard Sadoulet (1988). "Detection of Cosmic Dark Matter". In: *Annual Review of Nuclear and Particle Science* 28, pp. 751–807.

Randall, Lisa (2019). *The Worth of Physics Research*. URL: `https://www.nytimes.com/2019/02/01/opinion/letters/physics-research-collider-cern.html` (visited on 02/01/2019).

Riess, Adam G. et al. (1998). "Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant". In: *The Astronomical Journal* 116.3, pp. 1009–38.

Rubin, Vera C. and W. Kent Ford Jr. (1970). "Rotation of the Andromeda Nebula from a Spectroscopic Survey of Emission Regions". In: *Astrophysical Journal* 159, p. 379.

Rubin, Vera C., W. Kent Ford Jr., and Judith Rubin (1973). "A Curious Distribution Of Radial Velocities Of ScI Galaxies With 14.0<m<15.0". In: *The Astrophysical Journal* 183, pp. L111–15.

Sandage, Allan (1958). "Current Problems in the Extragalactic Distance Scale". In: *The Astrophysical Journal* 127, pp. 513–26.

Schneider, Mike D (2020). "What's the Problem with the Cosmological Constant ?" In: *Philosophy of Science* 87.January, pp. 1–20.

Sitter, Willem de (1933). "On the Expanding Universe and the Time-Scale". In: *Monthly Notices of the Royal Astronomical Society* 93.8, pp. 628–34.

Smeenk, Chris (2014). "Predictability crisis in early universe cosmology". In: *Studies in History and Philosophy of Science Part B: History and Philosophy of Modern Physics* 46, pp. 122–33.

Smeenk, Chris and Sarah C Gallagher (2020). "Validating the Universe in a Box". In: *Philosophy of Science* Forthcoming. arXiv: 2001.09101v1.

Staley, Kent (2004). "Robust Evidence and Secure Evidence Claims". In: *Philosophy of Science* 71.4, pp. 467–88.

Stanford, Kyle (2017). "Underdetermination of Scientific Theory". In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Winter 2017. Metaphysics Research Lab, Stanford University. URL: https://plato.stanford.edu/archives/win2017/entries/scientific-underdetermination/.

Stegenga, Jacob (2009). "Robustness, Discordance, and Relevance". In: *Philosophy of Science* 76.5, pp. 650–61.

Steinhauer, Jeff (2016). "Observation of Quantum Hawking Radiation and Its Entanglement in an Analogue Black Hole". In: *Nature Physics* 12.October, pp. 959–65.

Storms, Edmund (2007). *The Science of Low Energy Nuclear Reaction*. Singapore: World Scientific Publishing, p. 340.

Tegmark, Max (2003). "Parallel Universes". In: *Scientific American* May, pp. 40–51. arXiv: 0302131 [astro-ph].

Tolman, Richard C. (1934). *Relativity, Thermodynamics, and Cosmology*. Oxford: Clarendon Press, p. 501.

University of Utah (1989). *'Simple Experiment' Results in Sustained N-Fusion at Room Temperature for First Time*. Utah. URL: http://newenergytimes.com/v2/reports/UniversityOfUtahPressRelease.shtml (visited on 07/18/2019).

Unruh, William (1981). "Experimental Black-Hole Evaporation?" In: *Physical Review Letters* 46.21, pp. 1351–53.

— (2008). "Dumb Holes: Analogues for Black Holes". In: *Philosophical Transactions of the Royal Society* 366.June, pp. 2905–13.

— (2014). "Has Hawking Radiation Been Measured?" In: *Foundations of Physics* 44, pp. 532–45.

Unruh, William and Ralf Schützhold (2012). "Hawking Radiation From "Phase Horizons" in Laser Filaments?" In: *Physical Review D* 86, p. 064006.

Wasserman, Ira (1986). "Possibility of Detecting Heavy Neutral Fermions in the Galaxy". In: *Physical Review D* 33.8, pp. 2071–78.

Weinberg, Steven (2008). *Cosmology*. Oxford: Oxford University Press, p. 616.

Weinfurtner, Silke et al. (2010). "Measurement of Stimulated Hawking Emission in an Analogue System". In: *Physical Review Letters* 106, p. 021302.

Weinstein, Steven (2006). "Anthropic Reasoning in Multiverse Cosmology and String Theory". In: *Classical and Quantum Gravity* 23, pp. 4231–36.

Weisberg, Michael (2013). *Simulation and Similarity: Using Models to Understand the World*. Oxford: Oxford University Press, p. 224.

Weisberg, Michael et al. (2018). "The Dark Galaxy Hypothesis". In: *Philosophy of Science* 85.December, pp. 1204–15.

White, Simon D. M. and Martin J. Rees (1978). "Core Condensation in Heavy Halos: A Two-Stage Theory for Galaxy Formation and Clustering". In: *Monthly Notices of the Royal Astronomical Society* 183.3, pp. 341–58.

Wietfeldt, Fred E and Geoffrey L Greene (2011). "Colloquium: The Neutron Lifetime". In: *Reviews of Modern Physics* 83.October- December, pp. 1173–92.

Wimsatt, William C (2007). *Re-Engineering Philosophy for Limited Beings*. Cambridge, MA: Harvard University Press, p. 472.

Winsberg, Eric (2009). "A Tale of Two Methods". In: *Synthese* 169.3, pp. 575–92.

Witze, Alexandra (2019). "Physicists Close In on Neutron Puzzle". In: *Nature* 568, pp. 442–43.

Woodward, Jim (2006). "Some Varieties of Robustness". In: *Journal of Economic Methodology* 13.2, pp. 219–40.

Wylie, Alison (1999). "Rethinking Unity as a "Working Hypothesis" for Philosophy of Science: How Archaeologists Exploit the Disunities of Science". In: *Perspectives on Science* 7.3, pp. 293–317.

Yue, A. T et al. (2013). "Improved Determination of the Neutron Lifetime". In: *Physical Review Letters* 111, p. 222501.

Zel'dovich, Ya B (1965). "Survey of Modern Cosmology". In: *Advances in Astronomy and Astrophysics* 3, pp. 241–379.

Zwicky, Fritz (2009). "Republication of: The Redshift of Extragalactic Nebulae". In: *General Relativity and Gravitation* 41.1, pp. 207–24.