A Priori Entailment and the Metaphysics of Science

Kelvin McQueen

A thesis submitted for the degree of

Doctor of Philosophy of

The Australian National University

July 2013
Statement

This thesis is solely the work of its author. No part of it has previously been submitted for any degree, or is currently being submitted for any other degree. To the best of my knowledge, any help received in preparing this thesis, and all sources used, have been duly acknowledged.

I acknowledge and celebrate the First Australians on whose traditional lands I have lived and worked while writing this thesis, on whose traditional lands the Australian National University is located, and whose cultures are among the oldest continuing cultures.

Kelvin McQueen

July 2013
Acknowledgments

I thank the Chair of my supervisory panel, David Chalmers. It was while writing my master’s thesis back home in Dunedin that I gained respect for his contributions to philosophy. To then go on to work with him here in Canberra has been a great honour. His contributions to his students deserve equal respect. I thank him for his tireless critical feedback on my ideas and writing, his patience, and his generosity. I could not have asked for a better supervisor.

I thank the other members of my panel, Daniel Nolan, Jonathan Schaffer, Wolfgang Schwarz, and Daniel Stoljar. Each of them generously gave me their time whenever I needed them; they each brought their own critical perspective to this project, and challenged my thinking in a variety of ways. I am very grateful to them all.

To achieve what I wanted to achieve in this project I needed to learn a lot of physics. I could not have done this without the assistance of two friends who (perhaps unintentionally) acted as unofficial physics tutors. I thank Jesse Robertson—a philosopher in a physicist’s body despite what he may say—who helped develop my knowledge and appreciation of classical mechanics at a critical time. I also thank Michael Simpson for all the lunches—often at short notice, and for helping me think critically about quantum mechanics.

I thank the ANU philosophy department more generally, for providing an outstanding academic environment, and for providing me with many opportunities. I had the good fortune to visit Oxford University for five months, and to visit many other places for conferences and workshops. I was able to convene a reading group on quantum mechanics that helped me to build bridges with the physics department. I was introduced to a wide variety of philosophical topics through the department’s foundations seminars. Daniel Stoljar’s inaugural foundations seminar on physicalism was particularly influential. And then there were all those legendary games of soccer.

For helpful discussion and correspondence over the years I am grateful to Rachael Briggs, Adrian Currie, Zoe Drayson, Edward Elliott, Hilary Greaves, Alan Hájek, Frank Jackson, Dan Korman, Jano L, John Maier, Dan Marshall, Tim Maudlin, Angela Mendelovici, Gabriel Rabin, Raul Saucedo, Craig Savage, Laura Schroeter, Jon Simon, Alex Skiles, Stuart Szigeti, David Wallace, Timothy Williamson, Mark Wilson, and Alastair Wilson.

A special thanks must also go to the other members of the group of philosophers who will always be known as the Philosophical Incredulous Stair Society. John Cusbert, Jonathan Farrell, Jo Lau, Leon Leontyev, and Brian Rabern; you guys are inspirational.

Finally, I would like to thank my family, Christine, Jim, and Katie, for their continual support throughout my life. During my time in Canberra it was always a pleasure to come back home each Christmas, a trip that always left me invigorated for the year to come.
Abstract

In this dissertation I develop a framework for evaluating theories of fundamental reality and a related framework for evaluating reductive explanations. The former is the A Priori Entailment thesis (AET), which states that all truths are a priori deducible from the fundamental truths. The latter is the Reduction Entailment thesis (RET), which states that a successful and complete reductive explanation requires that the explanandum is a priori deducible from the explanans. After defending AET/RET I use them to resolve disputes in quantum metaphysics.

The dissertation is split into four chapters. The first chapter, Evaluating the A Priori Entailment Thesis, motivates underlying semantic and epistemological theses. It then defines a procedure for evaluating AET: a priori entailment expansions. These formulate problem cases into complex conditionals (of the form ‘If [fundamental truths] then [non-fundamental truths]’) and breaks them down into a number of simpler conditionals whose epistemic statuses are easier to evaluate. A priori entailment expansions are structured to guarantee that if the expansion conditionals are a priori then the expanded conditional is too.

Chapter two, A Priori Entailment as a Constraint on Classical Physical Theories, constructs a number of entailment expansions showing that AET is true of worlds described by classical physics. Each expansion conditional in each entailment expansion is a case study in itself, with the potential to support or undermine AET. The primary case study shows that mass additivity is an a priori consequence of Newton’s fundamental laws. I go on to argue that a particular metaphysical thesis explains these results. This metaphysical thesis strongly suggests that AET holds for any adequate fundamental theory, so AET must be true of the actual world.

Chapter three, Evaluating the Reduction Entailment Thesis, defends RET. I compare the mass additivity expansion to the scientific explanation of mass additivity in physics. I show that the expansion can be seen as the result of removing subtle simplification-induced explanatory gaps exhibited by the scientific explanation. I use this to argue that a priori entailment is necessary for reductive explanation, as it is necessary for the explanatory gaps to be innocuous. I develop a method for systematically converting any reductive explanation into an a priori entailment expansion. I extend the account to inhomogeneous reductive explanations. I then relate RET to the reduction of consciousness and the mind-body problem.

The final chapter, A Priori Entailment as a Constraint on Quantum Theories, assumes AET/RET and is the first detailed attempt at apply AET/RET to solutions to the quantum measurement problem. Such solutions are criticised for failing to reductively explain the manifest world. I evaluate such claims in the context of dynamical collapse theories. I object to the idea that such theories cannot explain why the manifest world appears three-dimensional. However, I provide new reasons for thinking that such theories cannot explain the existence of ordinary objects. In particular, I use AET/RET to show that dynamical collapse theories are empirically false.
Contents

1. Evaluating the A Priori Entailment Thesis
   1.1. Naturalistic Versus Conceptual Metaphysics 6
   1.2. Distinguishing A Priori Truths 16
   1.3. The Methodological Problem 22
   1.4. The Scale-Translation Problem 24
   1.5. A Priori Entailment Expansions 30

2. A Priori Entailment as a Constraint on Classical Physical Theories
   2.1. Mass Additivity 35
   2.2. The Concept of Mass 37
   2.3. Reduction of Mass Additivity 44
   2.4. Inertial Mass 51
   2.5. Gravitational Mass 60
   2.6. Gravitational Mass Density 67
   2.7. Charge 70
   2.8. Mereology 72
   2.9. Beyond Classical Physics 79

3. Evaluating the Reduction Entailment Thesis
   3.1. The Objection from Real Reductions 83
   3.2. Reduction of Conductivity 91
   3.3. Inhomogeneous Reductions 95
   3.4. Reduction of Consciousness 109

4. A Priori Entailment as a Constraint on Quantum Theories
   4.1. Measurement and Reality 117
   4.2. Dimensionality 120
   4.3. Dynamical Collapse Theories 129
   4.4. Quantum Metaphysics 144

References 148
Evaluating the A Priori Entailment Thesis

1.1 Naturalistic Versus Conceptual Metaphysics

Metaphysics is an enquiry into the nature of reality: what exists in reality and how reality is structured. The metaphysics of science differs from other forms of metaphysics by the methods it uses. The metaphysics of science attempts to understand the nature of reality using the sciences—their discoveries, explanations, theories, and concepts. The background assumption is that the sciences provide our best insight into the nature of reality, so metaphysics is best pursued by studying science itself.

One may distinguish two forms that the metaphysics of science can take. On one extreme there is naturalistic metaphysics, which rejects the traditional notions of the a priori and the analytic entirely, and treats the metaphysics of science as autonomous from conceptual analysis. On the other extreme, there is conceptual metaphysics, which treats the notion of the a priori and/or the analytic as indispensable, and therefore treats the metaphysics of science as requiring conceptual analysis at some level.

It is commonly thought that the metaphysics of science is synonymous with naturalistic metaphysics. For example, in the official definition of the metaphysics of science, The Society for the Metaphysics of Science states that “The metaphysics of science is neither transcendental nor aprioristic since it takes its foundation in the sciences”.¹ But this presupposes that a priori knowledge is irrelevant to the sciences. If, for example, scientific explanations presuppose a priori knowledge in significant ways, then the metaphysics of science should reject naturalistic metaphysics. Another reason why the metaphysics of science is often identified with naturalistic metaphysics is that conceptual metaphysicians typically do not examine the sciences in significant detail. In particular, conceptual metaphysicians develop abstract frameworks for examining metaphysical issues with little regard for whether such frameworks apply to successful scientific explanations and theories. I aim to remedy this.

¹ See https://sites.google.com/site/socmetsci/metaphysics-of-science (accessed 23/06/13).
In this section I discuss two scientifically motivated concepts crucial to metaphysical enquiry: \textit{grounding} and \textit{fundamentality}. I then discuss influential theories of grounding and fundamentality that have been proposed by conceptual metaphysicians. I build on these theories and formulate two theses that form the core of conceptual metaphysics as I propose to understand it: the A Priori Entailment Thesis and the Reduction Entailment Thesis. I then consider influential objections to these ideas proposed by naturalistic metaphysicians, before summarising the intended contribution of my dissertation to this debate.

An influential idea in metaphysics is that facts typically depend on (or obtain in virtue of) other facts. A related idea is that there is a special class of facts— the fundamental facts—on which all other facts depend, and which do not themselves depend on any further facts. The relevant notion of dependence in these ideas is neither causal nor historical. Perhaps future facts causally depend on past facts and perhaps all facts causally depend on facts about the initial condition of the universe. But this is not what is at issue. The relevant notion of dependence concerns a kind of constitutive determination dubbed \textit{grounding}.\textsuperscript{2}

The idea that facts are typically grounded in other facts is important to the study of scientific explanation. For example, to understand why a particular cup broke, we might ask: in virtue of what did the cup break? Read causally, the answer might be: in virtue of it being knocked off the table. But read as a grounding question, the answer might appeal to the breaking of relatively weak intermolecular bonds among constituent molecules. If this latter answer is correct, then the fact that the cup broke is grounded in facts about the behaviour of certain intermolecular bonds. Knowing this enables us to understand why things are brittle. As another example, if we want to understand conscious experiences, we might ask: in virtue of what am I experiencing blue rather than green right now? Read causally, the answer might appeal to the fact that I am looking to the sky above rather than to the field below. But read as a grounding question, the answer might appeal to underlying brain states. If this latter answer is correct, then certain facts about my experience are grounded in certain facts about my brain. In some cases, we might not expect answers to grounding questions. For example, if we were to ask the following question (a question that makes little sense under a causal reading): in virtue of what are events ordered in time? Perhaps the answer is: they just are. In that case, the fact that there is time is not grounded in any further facts and so the fact that there is time is \textit{fundamental}.

The idea that there is a special class of fundamental facts which ground all other facts is partly motivated by the idea that a branch of science—physics—routinely postulates a fundamental level of

\textsuperscript{2} Useful introductions to grounding can be found in Clark and Liggins (2012) and Trogdon (forthcoming). The notion was introduced by Fine (2001).
facts which it treats as its primary object of enquiry. Furthermore, many philosophers believe that experiential facts (facts about experience) are grounded in neurological facts, and that neurological facts are ultimately grounded in physical facts. If all non-physical facts are related to physical facts by some chain of grounding relations, and physical facts are not grounded by further facts, then the physical facts are the fundamental facts. The metaphor of “nothing over and aboveness” is sometimes used to motivate the general idea. The non-fundamental facts are nothing over and above the fundamental facts: fix the fundamental facts and you fix all facts; fix all of the facts that ground the fact that there is table, and one need do nothing more to ensure the existence of a table.

All sorts of questions can be raised about fundamentality and grounding. The question that I want to ask is: how can we determine which facts are the fundamental facts and which facts are merely grounded in the fundamental facts? Two schools of thought are conceptual metaphysics and naturalistic metaphysics. Conceptual metaphysics relies on conceptual analysis. Naturalistic metaphysics instead relies only on the details of certain scientific explanations and theories.

Conceptual metaphysics assumes that conceptual analysis substantively contribute to explanations that answer grounding questions. A paradigmatic example of conceptual metaphysics is what Frank Jackson calls serious metaphysics, which he defends in From Metaphysics to Ethics: A Defence of Conceptual Analysis (1998). Serious metaphysics attempts to give a complete account of reality in a limited vocabulary. This project gives rise to what Jackson calls location problems. Consider the problem of locating experience. Should experiences be located in our minimal yet complete account of reality by being explicitly mentioned by that account? Because Jackson is a physicalist (i.e. someone who thinks only physical facts are fundamental), Jackson does not want experiences to be explicitly mentioned in the minimal yet complete account. But because Jackson believes that experiences exist, he must define a sense in which facts about experience are implicit within the minimal yet complete account. Here Jackson introduces his entry by entailment thesis. The entry by entailment thesis states that even if experiences are not explicitly mentioned in the account, experiences can still be located in that account provided that experiences can be shown to be entailed (in the sense of necessary truth preservation) by that account. This raises the issue of how the minimal yet complete account can be shown to entail other facts. According to Jackson, a fact is entailed by the minimal yet complete account if and only if that fact follows a priori from that account. Roughly, A follows a priori from B if and only if knowledge of B and mere understanding of A puts one in a position to know A. Jackson uses philosophy of language to explain understanding and how it gives rise to a priori entailment. Thus, Jackson’s project is a project in conceptual metaphysics, which provides a means for distinguishing what’s fundamental from what’s merely grounded in what’s fundamental: The fundamental facts are those that are explicitly described by the minimal yet

---

3 As Callender (2001) writes, “Science has always gone for fundamentalism […] Newton, LaPlace, Maxwell, Clausius, Einstein, Witten, and company have always posited a fundamental level.”
complete account of reality while the non-fundamental facts are those that are a priori entailed by that account.\(^4\)

Another project in conceptual metaphysics is undertaken by David Chalmers in *The Conscious Mind* (1996). Chalmers defends a view similar to Jackson's and his goal is to show that phenomenal facts (a type of experiential fact specifically concerning what it is like to be in some conscious state) do not follow a priori from physical facts, or from any non-phenomenal facts. This is partly an attempt to refute physicalism: the idea that fundamental facts are physical facts and that phenomenal facts are grounded in physical facts. A key argument for this conclusion is the so-called *conceivability* argument. The first premise of this argument states that zombies—physical duplicates of ourselves that are not conscious—are conceivable. Here ‘conceivable’ means ‘cannot be ruled out a priori’. Chalmers argues from the conceivability of zombies to the claim that zombies are metaphysically possible, and then to the claim that physicalism is false. Part of the argument for the conceivability premise is that there is no conceptual analysis of the concept of consciousness in physical or functional terms. This is a clear case of conceptual metaphysics at work.

In *Constructing the World* (2012) Chalmers builds on these ideas, by defending the claim that all truths follow a priori from a small class of truths. He briefly considers the *Fundamental Scrutability thesis*, which says that “all truths are scrutable from metaphysically fundamental truths” (2012: 8.6).

‘Scrubtable from’ is a technical phrase which roughly means ‘knowable from’. Chalmers argues for a qualified version of the thesis which states that all truths are a priori knowable from the fundamental truths and a “that’s all” clause and certain indexical truths. The “that’s all” clause says “and that is all that is fundamental”. To understand why this "that's all" clause is required, assume that all fundamental truths are physical truths. Given the physical truths, one cannot a priori infer that there are no ghosts because one cannot a priori infer that there are no further fundamental facts that a priori entail ghosts. The “that’s all” clause therefore ensures that negative facts such as ‘there are no ghosts’ are a priori derivable from the fundamental truths. I will discuss “that’s all” clauses in more detail later (section 1.4). The appeal to indexical truths in the qualified version of Fundamental Scrutability is more problematic. Chalmers admits that indexical truths are both non-fundamental and not a priori derivable from fundamental truths. So indexical truths look like counterexamples to Fundamental Scrutability. Throughout chapter two, the worlds I consider when evaluating AET do not satisfy any indexical propositions because they do not contain self-locating agents. So I will leave this issue for now and return to it in section 3.4.

I primarily focus on two theses of conceptual metaphysics. One is similar to Fundamental Scrutability. I call it the *A Priori Entailment Thesis* (or AET for short):

\(^4\) Jackson (2003) defends physicalism using this framework but previously used it to reject physicalism (1982).
A Priori Entailment Thesis (AET): The material conditional ‘If FTI then N’ is a priori, where ‘FTI’ is the conjunction of the fundamental truths, a “that’s all that is fundamental” clause, and certain indexical truths; and where ‘N’ is any non-fundamental truth.

It is possible to formulate AET as the view that the inference from FTI to N is an a priori inference. However, the formulation in terms of material conditionals simplifies a number of issues, and is a substantive thesis no matter what one’s view of ordinary indicative conditionals. I will often speak of there being an a priori entailment from P to Q, by which I mean that the material conditional ‘If P then Q’ is a priori.

In formulating AET it is important to be clear on what the objects of a priori knowledge are, and what truths are. Candidates include propositions, thoughts, beliefs, utterances, sentence-types, and sentence-tokens. It is possible to define apriority so that it applies to any of these things and so it is possible to frame AET in a number of ways. Here I treat propositions as primary, where a proposition is what is said by a sentence. A proposition is what we refer to when we say that two distinct sentences (e.g. an English sentence and its French translation) say the same thing.

I choose propositions because they make my exposition simpler. However, the downside is that the nature of propositions is contested. For example, some philosophers are eliminativists: they don’t like the seemingly abstract nature of propositions and so prefer to only posit concrete entities such as individual utterances and thoughts. Here I wish to emphasise that propositions are simply an expository device. If they don’t exist, then an a priori entailment from ‘Jones is a bachelor’ to ‘Jones is unmarried’ is not an a priori entailment between propositions but is an a priori entailment between instances of some other type of entity. Characterising the nature of this type of entity is a difficult issue, but not an issue I wish to focus on here.

Philosophers who posit propositions may still be concerned by the fact that different theories of propositions sometimes yield different results for what follows a priori from what. For example, a minimal conception of propositions identifies propositions with sets of worlds. On this view, simply inserting a necessary mathematical proposition into the fundamental base thereby inserts all necessary propositions such as the proposition that water is H₂O (assuming it’s necessary). But this would not result from a Fregean view of propositions which treats propositions as more structured entities that reflect cognitive significance. My response to this concern is similar to the response to the eliminativist. Propositions are for me an expository device. When I argue that one proposition a priori entails another I argue that the entailment exhibits paradigmatic features that are defining of apriority (section 1.1). If such arguments are inconsistent with a theory of propositions then (if the arguments work) the advocate of that theory must find some other entities that exhibit the relevant entailment.
Chalmers (2012: 2.2 & E3) defends the idea that a small class of truths a priori entail all truths using an alternative framework based on sentences. This is partly because Chalmers has different aims, such as using the framework to defend a certain theory of propositions. Given such aims, presupposing a theory of propositions, or framing things in terms of propositions while staying neutral on their nature, is less of an option. The appeal to sentences runs into the difficulty of making sense of the idea that we know sentences (as opposed to what sentences say). I am not trying to defend a view of propositions and so freely use them as an expository device. Chalmers’ project plays the useful role of showing how AET can be formulated in a variety of ways to avoid controversies about propositions.\(^5\)

Some conceptual metaphysicians appeal not to a priori entailment, but to the closely related notion of \textit{analytic entailment}. For example, Amie Thomasson in \textit{Ordinary Objects} (2007) is concerned with debates over the existence of composite macroscopic objects such as trees and tables and whether they can be shown to be grounded in microphysics. In these debates all parties agree that there are particles arranged table-wise, but there is disagreement over whether there are tables. Thomasson argues that we can only refer to things like tables if the semantics of terms like ‘table’ are constituted by associated \textit{application conditions}. And she argues that the application conditions we in fact associate with such terms yield analytic entailments going from sentences such as ‘there are particles arranged table-wise’ to sentences such as ‘there are tables’. Thus, Thomasson argues that in virtue of the analytic entailments, all parties are committed to the existence of composite macroscopic objects. She then gives a deflationary analysis of the ontological questions raised in the debates: the notion of ‘composite object’ invoked in the debate lacks application conditions and is therefore semantically defective. I am sympathetic with Thomasson’s project (see section 2.8) and believe that a defensible analogue of AET could have equally well been defined in terms of analytic entailment. I choose a priori entailment instead for reasons that may also have motivated Frank Jackson (2003: 174-175): many accept Thomasson’s arguments that associated application conditions fix the reference of terms. But they deny that these associated conditions give the meanings of these terms and thereby deny the analyticities. But whether or not propositions expressing the application conditions are analytic, it is still plausible that they are a priori.\(^6\)

Anyone sympathetic to naturalistic metaphysics will wonder what a priori and/or analytic entailments have to do with real explanations. Let \textit{reductive explanation} refer to whatever type(s) of explanation in science answer grounding questions. Conceptual metaphysicians argue that reductive explanations

\(^5\) Other relevant issues that I will neglect, but which are discussed in detail in Chalmers (2012) are issues relating to idealizations (2.7, 2.8, excursus 5, excursus 6), the possibility of unknowable truths (excursus 1, excursus 4), and applications to areas of philosophy other than metaphysics and the philosophy of science (excursus 11, excursus 15). While these are important issues, I do not have space to discuss them.

\(^6\) Schaffer (2009: section 1) criticises Thomasson’s argument by suggesting that the relevant constraints on reference determination don’t yield substantive analytic entailments because the constraints are better regarded as \textit{metasemantics} rather than \textit{semantics} proper. But whether or not the constraints yield analytic entailments for this reason, it is plausible that the constraints yield a priori entailments.
rely on, or presuppose these entailments in some way. For example, Brie Gertler (2002) argues that evidence for a reductive explanation must be deemed as such by the concept of the thing being reduced. Furthermore, she argues that one is justified in accepting a reductive explanation only if one’s evidence for the reduction would be deemed as evidence by an analysis of one’s concept. Gertler calls this the “Conceptual Basis of Justification”. She uses this to argue that the debate over physicalism turns on conceptual facts, thus bolstering the conceivability arguments against physicalism.\(^7\)

Chalmers and Jackson (2001) relate a priori entailment to reductive explanation. After arguing that ordinary macroscopic truths follow a priori from microphysical truths, Chalmers and Jackson relate the reductive explanation of macroscopic truths in terms of microphysical truths, to a priori entailment. They argue that the relevant microphysical and macroscopic phenomena are both epistemically contingent “in that they involve the actualization of just one of a host of coherent epistemic possibilities” (2001: 351). (An epistemic possibility is defined as “a specific way the actual world might turn out to be, for all one can know a priori” (2001: 324).) For Chalmers and Jackson, where a priori entailment is present, the epistemic contingency in the macroscopic phenomena is reduced to the epistemic contingency in the microphysical phenomena: there is no further epistemic contingency in the connection.

The contribution of a priori entailment to reductive explanation is something I aim to get clear on. And so, as well as evaluating the A Priori Entailment Thesis (AET), I also aim to evaluate an application of AET to reductive explanation. I formulate the thesis as follows:

**Reduction Entailment Thesis (RET):** For any proposed reductive explanation, if the explanans does not a priori entail the explanandum, then the explanation is either incomplete or defective.

I discuss RET in detail in chapter 3. AET and RET together constitute the core of the conceptual metaphysics of science as I understand it. The above illustrations of contemporary conceptual metaphysics are by no means exhaustive. They are specific examples of conceptual metaphysics that have influenced me the most.\(^8\)

In contrast to conceptual metaphysics is naturalistic metaphysics. Naturalistic metaphysics assumes that science—unaided by conceptual analysis or a priori justification—offers the kind of explanation that answers grounding questions. Naturalistic metaphysicians do not accept that conceptual analysis

\(^7\) As Gertler’s (1999, 2001, 2006, 2007) other contributions to the debate make clear, she follows Chalmers and concludes on conceptual grounds that phenomenal facts are fundamental.

or a priori justification can add anything of interest. In fact, naturalistic metaphysicians are sceptical of conceptual metaphysics, believing it to be in conflict with contemporary philosophy of science.

Some naturalistic metaphysicians argue that conceptual metaphysicians have failed to keep up with developments in the philosophy of science and are assuming views that philosophy of science has rendered outdated. Others argue that there are specific aspects of how scientific explanations generally work that are inconsistent with AET/RET in particular. Still others think AET/RET is refuted by specific cases of scientific explanation. These are the kinds of challenge I aim to overcome in this dissertation. I will briefly illustrate each of these three challenges in turn.

Steven Horst, reflecting on the state of contemporary philosophy of mind, has this to say:

“In philosophy of science, the aprioristic normative agenda of the Positivists has been abandoned in favor of approaches that study the various methods and models of individual sciences, and the prevailing view is that the special sciences are autonomous and not in need of vindication by proving their reducibility to physics. The earlier optimism that the special sciences would prove reducible to physics has turned out to be largely unfounded. Reductions, in the relevant sense of that word, have proven few and far between, not only in the human sciences but in the physical sciences as well. And yet philosophy of mind has continued to labor under the yoke of an outdated philosophy of science. Indeed, it might not be an overstatement to say that turn-of-the-millennium philosophy of mind is one of the last bastions of 1950s philosophy of science.”

(Horst 2007: 47-48)

Horst’s notion of reduction refers to explanations that require conceptual adequacy, which is something that a priori entailment and analytic entailment are intended to achieve. In particular, “An explanation of A in terms of B is conceptually adequate just in case the conceptual content of B is sufficiently rich to generate that of A without the addition of anything fundamentally new” (Horst 2007:46). Horst is assuming naturalist metaphysics—that the best way to answer grounding questions (e.g. the question of conscious experience) is whatever method is employed by our best science. And because philosophy of science has supposedly shown that our best sciences do not answer grounding questions by appeal to anything like a priori entailment or analytic entailment, conceptual metaphysics is outdated.

Ausonio Marras (2005) appeals to a general conception of how real reductive explanations work in science (where ‘reductive explanation’ is not understood in Horst’s restricted sense, but in terms of whatever in science answers grounding questions). Marras uses this to criticise versions of conceptual metaphysics that rely on a priori entailment:
“A number of philosophers […] have claimed that there are conceptual grounds sufficient for ruling out the possibility of a reductive explanation of phenomenal consciousness. Their claim assumes a functional model of reduction […] which requires a priori entailment from the facts in the reduction base to the phenomena to be explained. […] This is an unreasonable requirement—a requirement that no reductive explanation in science should be expected to satisfy. […] The question whether consciousness is reductively explainable […] is a fundamentally empirical question, not one that can be settled on conceptual grounds alone.”

(Marras 2005: 335)

What Marras ultimately argues, is that conceptual metaphysics requires that bridge laws follow a priori from fundamental truths, whereas in science, bridge laws are inductively inferred. Marras concludes that his “account of how bridge laws are ‘inductively derived’ from a comparison of two levels of facts […] flatly contradicts [the] claim that bridge laws are ‘logically supervenient on the low-level facts’, or that they are ‘entailed a priori’ by the latter” (Marras 2005: 342).

Finally, Max Kistler (2006) considers cases of micro-macro reduction in condensed matter physics which involve the reductive explanation of a macro-property described by a macro-equation in terms of micro-properties described by certain micro-equations. In particular, he notes that (i) many such reductions involve showing that solutions to macro-equations are approximations to solutions to the relevant micro-equations and that (ii) some of these reductions essentially require irreducibly inter-level “solvability conditions”. The moral that Kistler draws is as follows:

“Reduction cannot, as is sometimes claimed, be achieved by a priori bottom-up derivation. […] Some input from above (in other words, from knowledge of the macro-level) is needed to constrain the derivation. […] Reduction can be achieved with the help of an “inter-level” device (in the form of the solvability condition), which cannot itself be logically reduced in terms of purely microscopic constraints. This is an important result. It allows to refute such claims as Chalmers and Jackson’s (2001) that the information in a complete microphysical description of the world suffices to deduce all macroscopic truths.”

(Kistler 2006: 350-351)

I have sympathy for these concerns. There is some truth in the claim that conceptual metaphysicians have not kept up with the details of how scientific explanations work. In fact, as far as I can tell, no conceptual metaphysician has examined a reductive explanation in detail and shown that it exhibits a priori entailment. Accordingly, the above complaints have gone unanswered.

But in so far as conceptual metaphysicians have not kept up with how scientific explanations work, naturalistic metaphysicians have equally failed to keep up with developments in the philosophy of
language, particularly the mechanics of a priori entailment. After all, there is not much resemblance between the positivist project and contemporary conceptual metaphysics, and so Horst’s claim that contemporary philosophy of mind is the last bastion of 1950’s philosophy of science is tenuous.\(^9\) Furthermore, Marras’ idea assumes that conceptual metaphysics makes a particular prediction about the form of real scientific explanations. I will show that this assumption is unwarranted (section 3.1). Finally, Kistler’s argument that conceptual metaphysics is refuted by his case study underestimates the resources conceptual metaphysicists have available to deal with such problematic cases (section 3.2).

One of the primary aims of this dissertation is to use the metaphysics of science against naturalistic metaphysics. I look to real scientific theories and explanations for evidence of a version of conceptual metaphysics that has AET and RET at its core. I defend AET/RET by appeal to three arguments. The primary argument is an inductive one: I show that the frameworks are supported by a number of case studies in classical physics (chapter 2). The second argument claims that these results suggest an independently plausible metaphysical thesis about the nature of fundamentality (sections 2.9 and 3.4). If this thesis is right then these results are independent of classical physics such that the frameworks should hold for all physical theories. The third argument claims that these frameworks enable progress in the metaphysics of science. In particular, I use them to illuminate why dynamical collapse theories in quantum mechanics are empirically false, contrary to the received view (chapter 4). The frameworks make essential use of notions of traditional aprioristic analytic philosophy. The dissertation therefore doubles as a challenge to the dominant trend in contemporary philosophy of science that opposes such notions.

I focus in particular on physical theories and physical explanations. One reason for this is that it simplifies the task: if one were to focus on theories and explanations more generally, then the task becomes difficult due to the variety of theories and explanations that the various sciences have to offer. Physics is also the natural starting point when evaluating AET. That’s because of the connection between fundamental theories and physical theories. Known physical theories enable us to give complete fundamental descriptions of certain possible worlds. For example, Newtonian mechanics enables us to give complete fundamental descriptions of simple worlds in which Newtonian mechanics is true. As we shall see, this will prove very useful for evaluating AET.

\(^9\) For a comparison of the views of the positivists and a contemporary version of conceptual metaphysics see Chalmers (2012: sections 1, 3.1, 5, 8, E1, E10, E12, E15). For an extensive critical review of Horst’s book see Witmer (2008).
1.2 Distinguishing A Priori Truths

Several important arguments in this dissertation rely on premises stating that certain truths are a priori. However, the very idea of apriority is controversial and so is any claim stating that some particular truth is a priori.\(^\text{10}\) So right away, something needs to be said about how we should understand apriority and how we should go about classifying truths as being a priori or otherwise.

It would be inappropriate to develop a general theory of the a priori\(^\text{11}\) for at least two main reasons. Firstly, doing this properly would require a whole dissertation and would therefore overshadow the primary goal of the dissertation—develop AET/RET so it can be usefully applied in the metaphysics of science. Secondly, developing a theory of the a priori is unnecessary if the goal is to respond to naturalist metaphysics based objections to AET/RET. None of the objections I consider are based on general scepticism of the a priori. Rather than appealing to developed epistemic or semantic theories, the objections appeal to features of scientific explanations and theories. The objections grant (for the sake of argument) that propositions in mathematics, logic, and perhaps other domains are a priori. They instead deny that the class of a priori truths are broad enough to be useful to the metaphysics of science.

With this in mind, I will simply specify two plausible heuristics for distinguishing a priori truths. I use these heuristics throughout the thesis to argue that we have evidence to think that the propositions that need to be a priori for AET/RET to be defensible, are in fact a priori.

A standard way of defining apriority is as follows:

\[
\text{A priori justification} \equiv \text{justification that is independent of evidence from experience.}
\]

\[
\text{A priori knowledge} \equiv \text{knowledge one has with a priori justification.}
\]

\[
\text{A priori truth} \equiv \text{truth one can have a priori knowledge of.}
\]

The definition of a priori justification does not appeal to total independence from experience but only to independence from \textit{evidence from experience} or equivalently \textit{empirical evidence}. Justification for believing some proposition requires that one understand the proposition, which requires the possession of concepts, which typically requires experience. For example, we plausibly know that \textit{if the sky is blue then the sky is blue} with justification that is independent of evidence from experience. However, experiences of colours and skies enabled us to possess the concepts SKY and BLUE, and hence enabled us to understand the proposition. Experience therefore plays an \textit{enabling role} in our

\(^{10}\) Much scepticism is due to Quine (1951) although see Grice and Strawson (1956) for an underappreciated response.

\(^{11}\) For a detailed annotated bibliography on theories of the a priori see Casullo (2012a).
having a priori justified beliefs. In the case of empirically justified (a posteriori) beliefs, experience plays a further, evidential role.\textsuperscript{12}

According to the standard definition, an a priori truth is a truth that can (in principle) be known independently of empirical evidence. This definition allows that such truths can also be known empirically. For example, if perception enables you to hear a renowned mathematician assert that some mathematical theorem is true, then arguably this is empirical knowledge of an a priori truth. This type of phenomenon is common in introductory logic courses. Consider some slightly more complex logical truths, such as De Morgan’s laws:

\[
\neg(P \land Q) = \neg P \lor \neg Q \\
\neg(P \lor Q) = \neg P \land \neg Q
\]

The first law states that the negation of the conjunction of two propositions is equivalent to the disjunction of the negations of those propositions. Consider a “lazy” logic student who sees the first law written out on the white board (with the variables replaced by example sentences), but does not think reflectively about what the sentence says. In particular, the student glances over the sentence, understands each individual word, and briefly checks the entire sentence for grammaticality, and hence, understands the proposition expressed by the sentence at some level. But because the student does not reflect on the proposition, the student does not come to believe it. Instead, the student might visually experience his logic teacher (who he trusts as an authority on the issue) confidently assert that De Morgan’s laws are true. The student might then come to believe, and indeed know, the truth of De Morgan’s laws, on this basis. This is empirical knowledge of an a priori truth. A priori truths (truths knowable through entirely a priori methods, such as the methods employed by the logic teacher) can sometimes be known on the basis of empirical evidence (such as testimony). Another possibility is the converse. A student might inadequately reflect on a logical falsehood and come to believe it, only to be told by her logic teacher that it is logically false. This illustrates the possibility of empirical defeaters of a priori justification.\textsuperscript{13}

In what follows I provide two heuristics for characterising a priori truths that each take the form: ‘If a proposition p is such that φ then this is evidence that p is a priori’ for some epistemic property φ. The epistemic properties appealed to by each heuristic are properties that are symptomatic of being a

\textsuperscript{12} Burge (1993) provides a useful discussion of this distinction. Williamson (2007: 165-9, 2012) objects to the theoretical usefulness of the concept of a priori knowledge, by arguing that experience can play an epistemic role that is neither purely enabling nor strictly evidential. Casullo (2012b: section 5) identifies a number of gaps in Williamson’s argument. Chalmers (2012: 194-97) argues that a more exhaustive characterisation of experience’s epistemic roles removes the objection.

\textsuperscript{13} Burge (1993, 1997) develops a theory of testimony which allows testimonial a priori knowledge and which would deem the two cases just described to be a priori justification for an a priori truth and an a priori defeater for a priori justification, respectively. See Christensen and Kornblith (1997), Malmgren (2006) and Casullo (2007) for critiques.
priori. I will sometimes refer to these epistemic properties as the *apriority symptoms*. Thus, think of apriority on the analogy of a disease (possessed by propositions) whose nature we are unsure of. We can still pick out instances of things that possess the disease. And we can do so by appeal to paradigmatic symptoms. Symptoms can sometimes be misleading, but can still yield evidence for the presence of a disease. Accordingly, if the heuristics are applicable to a proposition, this only yields evidence for the apriority of that proposition. But this will be sufficient when it comes to defending AET/RET against naturalistic metaphysics based objections, while still remaining relatively non-committal regarding the epistemology of the a priori. The two heuristics are as follows:

**Analysis Based Justification**: If analysis of the concepts in a proposition enables one to justify their belief in the proposition, then this is evidence that the proposition is a priori.

**Negation Inconceivability**: If the negation of a proposition is inconceivable then this is evidence that the proposition is a priori.

In what follows I discuss each heuristic in turn. I illustrate the applicability of each heuristic for propositions normally considered a priori, such as logical, mathematical, and definitional propositions. I then briefly discuss them in the context of propositions that play important roles in my defence of AET in chapter 2.

**Analysis Based Justification**

Analysis based justification is illustrated in the context of propositional logic. Thus, consider how one might argue that the proposition ‘If p then p or q’ is true. One might appeal to what one means by ‘or’, in terms of the disjunction truth table. That is, one might say that it is part of what one means by ‘or’ that a statement of the form ‘p or q’ is true if one of the disjuncts is true. Similar things can be said for first order logic. One might explain why ‘If a is F then there exists something that is F’ is true by appeal to what one means by the existential quantifier. That is, one might say that it is part of what one means by ‘there exists’ that a statement of the form ‘there exists something that is F’ is true, just in case there is something, call it some arbitrary name ‘a’, such that a is F. It is standard to consider these truth conditions as a part of the semantics of logical connectives and quantifiers. And these conditions hold no matter what the actual world is like. Therefore, one does not evaluate such logical truths by appeal to empirical facts about the actual world, but by appeal to the conditions under which the logical concepts apply, to any possible world—never mind which is actual.

Here I assume that such logical propositions are a priori, and note an important epistemic property of logical propositions: that belief in them is (or can be) justified using analysis based justification. This is not necessarily to say that such propositions are analytic, or true in virtue of meaning. Rather,

14 Field (1995-1996) argues that principles of classical logic are not only a priori, but empirically indefeasible.
it is to simply diagnose an important epistemic property of a priori truths, so that that property can be used as a heuristic for diagnosing other, perhaps more complex truths as being a priori. The heuristic should therefore be acceptable to conceptual metaphysicians such as Bonjour (1998) who believe that analyticity cannot explain the apriority of logical truths, and who instead postulate special faculties of rational insight.\textsuperscript{15} So if a proposition admits of analysis based justification, then this is evidence that it is a priori.

Williamson (2007) objects to conceptual metaphysics by objecting to what he calls the “understanding-assent link”. But Analysis Based Justification does not commit to the understanding-assent link. That is, it does not imply that understanding such a priori logical propositions entails that one will accept them. Since ‘understanding’ comes in degrees, there will be many cases in which one ‘understands’ such propositions without accepting them, as Williamson demonstrates. Analysis Based Justification requires not only that one understand the relevant propositions, but that one is in a position to analyse the components of the propositions, and uses such analyses to defend one’s acceptance of the propositions. If belief in a proposition can be rationally justified using analysis based justification, then this is evidence the proposition is a priori.\textsuperscript{16}

Analysis based justification is not confined to logic classrooms. A familiar phenomenon in debates involves one disputant defending a claim by saying ‘this is just what I mean by the term’. Thus, imagine a debate over whether the pope is a bachelor. One disputant states that the pope is clearly not a bachelor because he is not on the dating scene, the other disputant states that the pope is a bachelor because he is an unmarried man, and ‘unmarried man’ is all that this disputant means by ‘bachelor’. If we take both disputants to be competent speakers of English, we can disambiguate the notion ‘bachelor\textsubscript{1}’ used by the former disputant and ‘bachelor\textsubscript{2}’ used by the latter disputant and diagnose this is a merely verbal dispute. In that case, the proposition ‘all bachelors\textsubscript{2} are unmarried men’ is a priori: analysis of its concepts enables one to justify their belief in it.\textsuperscript{17}

There is no reason to think that this kind of analysis based justification does not extend to propositions whose key components are metaphysical concepts such as ‘composite object’. Consider metaphysical debates over when composition occurs. I accept unrestricted composition, which states that any objects compose a composite object. I defend my acceptance of unrestricted composition by stating that I use ‘composite’ to refer to any objects that I can (in principle) group together in thought and label with a name (‘composite object 1’) so that I can make useful statements using the name (‘composite object 1 exists and behaves in such and such ways’). Because of this analysis based

\textsuperscript{15} Compare Hale (2002) who argues that acceptance of propositions stating basic rules of inference constitute understanding of logical concepts, which explains why they cannot be rationally doubted.

\textsuperscript{16} Balcerak Jackson and Balcerak Jackson (2011) defend conceptual metaphysics against Williamson’s objection in a similar manner.

\textsuperscript{17} Chalmers (2011) argues that analysis of verbal disputes can be used to reconstruct a notion of analytic truth defined in dialectical terms.
defence of unrestricted composition, I take myself to be a priori justified in believing unrestricted composition. Of course this is a highly controversial issue. I develop this view on the epistemology of mereology in section 2.8 and consider several objections to it. It is worth noting that, as discussed in 2.8, AET/RET requires only the apriority of a claim weaker than unrestricted composition that metaphysicians of a wide variety can accept.

There is no reason to think that analysis based justification does not extend to propositions whose key components are theoretical concepts such as ‘mass’. Consider the proposition that *if an object has the natural property that disposes it to resist changes in motion given applied forces then the object has mass*. Arguably, one can defend one’s belief in this proposition, not by appeal to experiences of how the actual world happens to be, but by appeal to the standard inertial definition of ‘mass’. I discuss this definition and the concept of mass in section 2.2. In section 2.4 I say precisely what is required in the antecedent of the italicised conditional in order for the consequent to be a priori deducible.

**Negation Inconceivability**

The Analysis Based Justification heuristic suggests that sufficient reflection on the concepts within an a priori proposition can enable one to see that the proposition is true. If this is correct, then the negation of an a priori proposition will be difficult to make sense of. Indeed, the negations of a priori propositions are typically inconceivable, or seem incoherent. Negation Inconceivability can therefore be used as a more direct heuristic in itself. If it is not possible to conceive a situation in which the negation of a proposition is true, then no possible situation falsifies the proposition. And if no possible situation falsifies a proposition then it does not matter how the actual world is: however it is, it cannot falsify the proposition. This does not necessarily entail that the proposition is true: perhaps there is simply no fact of the matter, because no world makes the proposition true, either. However, when one finds that one cannot put the concepts in a proposition together to conceive of how its negation could be true, then one will often have an *intuition of incoherence*. For example, take the proposition ‘If Jones is a bachelor then Jones is unmarried’. One cannot conceive of a married bachelor. Furthermore, the very idea seems incoherent. And arguably, this seeming, this intuition of incoherence, justifies one in believing the proposition whose negation is inconceivable.\(^{18}\)

If one could conceive the negation of a proposition, one would presumably require some empirical input to ensure that the actual world does not correspond to what one is able to conceive. The proposition would then not be a priori. And so for a given proposition, applying the Negation Inconceivability heuristic requires that one thinks of a situation that makes sense, in which the proposition is false. One way of doing this is to consider whether some possible experiment could

---

\(^{18}\) It is common to distinguish intuition-based explanations of apriority from meaning-based explanations (see Hawthorne (2000) and Dogramaci (2012)). However, a priori justification based on incoherence intuitions suggests a hybrid explanation.
conceivably falsify the proposition. If no such falsifying experiment can be conceived, then we have evidence of apriority. Paradigmatic a priori truths exhibit this feature. For example, finding bachelors and asking them if they are unmarried is a defective empirical test for the claim that If Jones is a bachelor then he is unmarried. For if a subject sincerely claimed to be a married bachelor, we would not take that as evidence against the conditional, but as evidence that the subject does not understand the question. Similarly, there is no conceivable empirical test for \(1+1=2\). Putting one apple, then another apple, in one’s hands, then seeing if there are two apples, is another defective experimental strategy. For even if the laws of nature conspired in such a way so as to always generate a third apple, when a person places two apples in their hands, we would not then accept that we live in a world in which sometimes, \(1+1=3\).

The Negation Inconceivability heuristic applies to the propositions crucial to my defence of AET/RET. For example, while unrestricted composition is controversial, if one is clear on the concept of composition that one is employing, and clarifies it in the manner I have, then the negation of unrestricted composition should be inconceivable. After all, unrestricted composition is an a priori consequence of the application conditions I gave for the concept of a composite. Those that reject unrestricted composition must therefore think that the debate should be framed in terms of a less committal concept of a composite. I discuss this response in detail in section 2.8 and argue that it does not work.

Arguably, one cannot conceive of a composite being located in a region far removed from the regions in which its parts are located. Furthermore, it is hard to make sense of how a composite might have a velocity that is distinct from the velocities of its parts, when the velocities of its parts are identical. These propositions will play an important role in chapter two, where I will discuss their epistemic statuses in more detail.

If propositions that are not a priori, such as ‘water is H\(_2\)O’, seem necessary, then the Negation-Inconceivability heuristic might seem unreliable. Similarly if some a priori propositions seem contingent such as ‘Julius (if he exists) invented the zip’. In response we just need to distinguish different notions of necessity. The relevant sense in which a priori truths seem necessary is the sense in which their negations are inconceivable: one cannot conceive of a situation in which the negation is true as such situations seem incoherent. The negations of necessary a posteriori propositions are conceivable, while the negations of contingent a priori propositions are inconceivable. For example, it’s conceivable that the actual world is such that water isn’t H\(_2\)O: if tomorrow chemists tell us that water is not H\(_2\)O, we are not going to find their claim to be incomprehensible.\(^{19}\)

\(^{19}\) For a more formal discussion of this notion of epistemic necessity see Chalmers (2004) and Elliott, McQueen and Weber (2013).
One might worry that this notion of conceivability somehow presupposes the notion of apriority. However, we are all familiar with the activity of conceiving or imagining situations and we are all familiar with the incoherence of square circles and married bachelors. Such familiarity hardly assumes the a priori in any relevant sense: folk can make reasonable judgments about the incoherence or inconceivability of some proposition without ever coming across the philosophical notion of the a priori. Inconceivability might be best analysed in terms of what can be ruled out a priori, but this does not mean that our heuristic presupposes apriority in any question-begging way.

Our two heuristics can’t prove that a given proposition is a priori, they only provide evidence. But this is sufficient given our dialectical purposes: if one can show that non-fundamental truths can be derived from fundamental truths with inferences or premises that satisfy our heuristics, then the naturalistic metaphysician cannot simply appeal to features of real explanations and scientific theories to object to conceptual metaphysics. They must leave the domain of philosophy of science altogether and do epistemology or philosophy of language, and explain what other property the heuristics are tracking, if not apriority.20

1.3 The Methodological Problem

AET states that the material conditional ‘if FTI then N’ is a priori. However, we don’t, and perhaps can’t understand F, which raises a methodological problem:

Methodological Problem for AET:
(1) We cannot fully understand F.
(2) If we cannot fully understand F then we cannot know whether non-fundamental truths are a priori entailed by FTI.
(3) Therefore we cannot know whether non-fundamental truths are a priori entailed by FTI.

In response, here are three solutions which each deny (2):

20 Devitt (2005, 2011) attempts this using the idea that an individual’s beliefs form an interconnected web, such that the conjunction of all of one’s beliefs are confirmed or disconfirmed by empirical evidence (“confirmational holism”). Devitt argues that a belief’s property of “being in the interior of the web”—being among the beliefs held fixed as one revises other beliefs in response to empirical evidence—explains apriority symptoms. However, the idea is too underdeveloped to adequately explain negation inconceivability (Devitt 2005: 107-108), and is difficult to reconcile with real developments in mathematics and hence the metaphysics of science itself (Maddy (1992: 286-89)). See Bonjour (2005) for further critique.
Epistemological Optimism: We have reasonable candidate fundamental theories (e.g. quantum gravity theories), which give us a partial understanding of F, and we can evaluate AET by seeing whether these theories support AET.

Other Possible Worlds: If AET is true then either AET is necessarily true or AET is (at least) true in close possible worlds. This enables us to evaluate AET by evaluating it at close possible worlds, such as Newtonian worlds.

High-Level Entailment: AET suggests that a priori entailment relations correspond to grounding relations. Grounding relations obtain between the explananda and explanantia of high-level special science reductive explanations. Therefore we can evaluate AET by examining whether successful high-level reductive explanations exhibit a priori entailments.

I discuss all three strategies in what follows, before pursuing Other Possible Worlds in detail. Epistemological Optimism treats premise (2) as assuming an unwarranted pessimism. We have well-developed fundamental physical theories. We can describe the physical states of isolated regions of the universe using the vocabulary these theories provide. We can then analyse what follows a priori from these descriptions. If nothing interesting is a priori deducible from such descriptions then AET is in trouble, because AET would be at odds with modern physics. On the other hand, if interesting non-fundamental truths are deducible, then AET receives support from quantum gravity theories. Developing Epistemological Optimism would involve a very ambitious project which I leave for future research.

Chapter two develops Other Possible Worlds and provides support for AET by showing that it holds in Newtonian possible worlds. This assumes that Newtonian possible worlds are close possible worlds. For our purposes a world is close if the fact that AET holds in that world supports the truth of AET (in the actual world). The closeness of the world is therefore proportional to the extent to which the world supports AET if AET is true of that world. Newtonian worlds are close in the relevant sense. Any proposed fundamental theory must exhibit some structural isomorphism to the Newtonian theory it is replacing. For example, it must in some sense recover the success of Newtonian mechanics within its domain of validity. Our world being fundamentally structured in such a way so as to exhibit high-level structure that can be predicted, explained, and manipulated, when we describe it in Newtonian terms, is reason to think Newtonian worlds are close possible worlds. Furthermore, Newtonian physics is now best seen as a special science which is capable of providing reductive explanations within its (non-fundamental) domain of validity. Arguably, worlds that are describable (to a sufficient degree of accuracy) in the terms of our high-level sciences are close possible worlds.

21 For example the prominent theories of quantum gravity (Oriti 2009).
22 The structural isomorphism requirement is referred to as the principle of correspondence in the philosophy of physics literature (Albert 1994: 44).
Newtonian worlds are describable in terms of our high-level sciences. Therefore Newtonian possible worlds are close possible worlds.\(^{23}\)

Chapter three develops High-Level Entailment and in the process evaluates RET. The idea is that because AET postulates such an intimate connection between a priori entailment and being grounded in the fundamental facts, then if grounding relations described by special sciences exhibit a priori entailment, then there is an inductive argument to the conclusion that all grounding relations exhibit a priori entailment.

In the next section I consider an influential method for evaluating AET, which appeals to worlds fundamentally described by microphysics. I argue that the method faces a problem the scale translation problem. I then introduce a framework which will allow me to solve both the methodological problem and the scale translation problem.

1.4 The Scale Translation Problem

Impressed by the apparent ability of physics to provide a complete description of reality in minimal terms, conceptual metaphysicians typically treat the fundamental description of reality as a physical description ("physicalism"). The description is typically thought of loosely as a description of elementary particles related by forces, interacting in spacetime ("microphysicalism").\(^{24}\) Thus, AET is typically evaluated through the evaluation of AET*:

**A Priori Entailment Thesis* (AET\(^*\)):** Ordinary macroscopic truths follow a priori from physical truths.

AET\(^*\) states that ordinary macroscopic truths—scientifically non-mysterious truths about the manifest world that uncontroversially admit of in principle reductive explanation—follow a priori from physical truths. The thought behind AET\(^*\) is that we know from our best total science that fundamental truths are likely to consist (mostly if not entirely) in physical descriptions and we also know that some macroscopic descriptions are straightforward non-fundamental truths; and so, providing support for the more manageable AET\(^*\) will provide support for the more elusive AET.

AET\(^*\) has played a significant role in recent philosophy largely because the viability of an influential account of reductive explanation depends on it:

\(^{23}\) The idea that Newtonian mechanics is (now) a special science and that Newtonian forces are non-fundamental ‘higher-level’ theoretical postulates, is defended by Wilson (2007).

\(^{24}\) Pettit (1993) defines physicalism in terms of microphysicalism.
Reduction Entailment Thesis* (RET*): If macroscopic truth M can be reductively explained in physical terms then M is a priori entailed by physical truths.

Without AET*, RET* would not be plausible. For if truths that we have successfully reductively explained don’t follow a priori from fundamental physical truths then RET* would not be a reasonable constraint on reductive explanation. RET* is applied in debates about the explanation of phenomenal consciousness: if phenomenal truths can be reductively explained in physical terms then phenomenal truths are a priori entailed by physical truths. Assuming this thesis (or something resembling it) many philosophers argue that phenomenal truths do not follow a priori from physical truths and conclude that phenomenal consciousness is irreducible. Others argue that phenomenal truths do follow a priori from physical truths and conclude that phenomenal consciousness is in principle reducible. This is largely a sociological phenomenon, as RET applications could apply to just about any area of philosophy. Consider the quantum physics application: If there is no a priori entailment from the fundamental description provided by quantum theory T to macroscopic truth M then reductive explanation of M in terms of T is impossible. Thus, many important debates in philosophy depend on the plausibility or otherwise of AET*.

Perhaps the most influential defence of AET* can be found in Chalmers and Jackson (2001). There, the defence is split into two distinct arguments. The first argues for the a priori entailment of macrophysical truths from microphysical truths. The second argues for the a priori entailment of macroscopic truths from macrophysical truths (perhaps in conjunction with microphysical truths).

Why microphysical truths? The thought is that (some specific set of) microphysical truths are good candidates for fundamental physical truths. However, AET* is not committed to a fundamental ontology of microscopic entities. For example, if monism—the thesis that all fundamental properties are properties of the universe as a whole—is true, then our fundamental physical truths will reference those properties. I follow the AET literature and invoke descriptions of microphysics as my candidate fundamental physical truths, though I will add the fundamental laws of nature to them. In doing so I do not wish to prejudge the issue of whether laws are non-fundamental, and grounded in something more fundamental. Laws are arguably fundamental, and so for the purposes of defending AET, their inclusion into our candidate F should not be controversial. It is better to ask questions

---

26 For example, Braddon-Mitchell (2003), Haukioja (2008), Jackson (2003a), Kirk (2008), and Lewis (1994).
27 Chalmers and Jackson (2001: 316-20) add indexical truths and a “that’s all” clause to the microphysical truths. They also add phenomenal truths (hence arguing that ‘If PQTI then N’ is a priori). Joseph Levine (2010) argues for the need to defend AET without phenomenal truths. I respond in section 3.4.
29 Such as regularities of events (Lewis (1986), Schaffer (2008)) or symmetries (Stenger (2006)).
about the fundamentality and a priori entailment of laws after we have shown what follows a priori from a base description that includes them.

Microphysical, macrophysical and macroscopic descriptions are defined as follows:

**Microphysical description**: proposition(s) describing the fundamental properties and relations of non-composite (elementary) microphysical entities (excluding reference to properties such as ‘being part of a composite’), and the fundamental physical laws.

**Macrophysical description**: proposition(s) describing composite physical entities entirely in the language of physics.

**Macroscopic description**: proposition(s) describing composites of physical entities not given entirely in the language of physics.

Thus, two questions must be asked when evaluating AET*. Firstly, do macrophysical descriptions follow a priori from microphysical descriptions? Secondly, do macroscopic descriptions follow a priori from macrophysical (plus microphysical) descriptions? The emphasis on *descriptions* rather than truths is due to the working assumption that a priori entailments hold between false propositions. For example, if AET* is true of close possible worlds then it presumably predicts that macrophysical and macroscopic descriptions that are true in a Newtonian possible world, will follow a priori from the fundamental physical description of that world.

There is an a priori entailment from P to M if and only if the material conditional ‘if P then M’ is a priori. AET* thus predicts that there is a large range of (non-trivial) a priori instances of the following schemas:

**(AET-1)**: If [microphysical description] & T then [macrophysical description].

**(AET-2)**: If [[macrophysical description] & [[microphysical description] & T]] then [macroscopic description].

Here a microphysical description is a fundamental description of a close possible world and ‘T’ says “that’s all that’s fundamental”.

In trying to solve the AET Methodological Problem by treating our fundamental description as microphysical, a new problem emerges:

**Scale-Translation Problem for AET**: Inferring macrophysical truths from microphysical truths involves a change in scale, in particular, a transition from small-scale entities to large-scale entities. There is no reason to think that from a microphysical description, we can just "scale-up in thought" and deduce truths about large scale entities.
Chalmers and Jackson’s two-step defence is intended to help here. In particular, AET-1 concerns inferences from microphysical descriptions to large-scale physical descriptions, while AET-2 concerns inferences from these large-scale physical descriptions, to large-scale non-physical descriptions.

In defence of the claim that there are non-trivial instances of AET-1, Chalmers and Jackson appeal to the apriority of what I call *scale-translation bridge principles*. Bridge principles relate levels of description, and often enable inter-level (reductive) explanations. For example ‘water is composed of H₂O molecules’ relates macroscopic-level truths with molecular-level truths, enabling explanations of water’s macroscopic properties in terms of molecular physics. A *scale-translation* bridge principle is a general principle that tells us how a predicate applies at a mereologically higher level given how it applies at a mereologically lower level. Two examples are: if there are particles located at x₁ and x₂ then there is a composite located at the set of positions \{x₁, x₂\}; and if particle 1 has mass m₁ and particle 2 has mass m₂ then the composite composed of particles 1 and 2 has mass m₁+m₂ (the principle of mass additivity).

Consider what Chalmers and Jackson say about how macrophysical descriptions follow a priori from microphysical descriptions:

“The only worry might concern the status of bridging principles within physical vocabulary: for example, is it a priori that the mass of a complex system is the sum of the masses of its parts? If there are any concerns here, however, they can be bypassed by stipulating that the relevant physical principles are built into P.” [Where P is (or was) the microphysical description.]

(Chalmers and Jackson 2001: 331)

The thought is that there are scale-translation principles that are either a priori or not a priori. If they are a priori then there is no difficulty inferring macrophysical descriptions. If they are not then we just add them to the microphysical description. The latter suggestion is not promising however. If one added to the microphysical description a claim about composite mass then it is no longer a purely microphysical description. This doesn’t matter so long as our addition is plausibly fundamental. But composites and their masses are plausibly non-fundamental (in fact later I demonstrate that composites follow a priori from classical microphysics). Thus, adding mention of composites and a general principle about their masses is not a permissible addition to the fundamental microphysical description.30

---

30 For reasons why the parthood relation is an impermissible addition to the fundamental description, see Sider (2013). I argue in section 2.8 that composition principles follow a priori from microphysical descriptions. Sider claims to be defending nihilism, though only in light of his assertion that “we should reformulate nihilism as the
Let’s relate this to the use of AET* to motivate RET*. If we find a reducible truth (e.g. about how much mass a composite has), which does not follow a priori from microphysics, appending a bridge principle to the microphysical description so that the required entailment then goes through is not an option. One may achieve an entailment by doing this. But if one wants a viable theory of reduction then this is not the right conclusion to draw. To bring this out, consider the philosophy of mind analogy. Why should mass additivity get off the hook, but not, say, psychophysical bridge principles? Compare:

(Mass): If [microphysical description + scale-translation bridge principle] then composite object o has mass m.

(Consciousness): If [microphysical description + psycho-physical bridge principle] then organism o is phenomenally conscious.

For the AET*/RET* advocate, the question of whether consciousness is reducible is not the question of whether (Consciousness) is a priori. In that case, why would the question of whether macrophysical mass is reducible be decided by the apriority of (Mass)?

Bypassing the concern that scale-translation principles are not a priori by stipulating them into the entailment base is therefore not an option. The other option mentioned by Chalmers and Jackson is that the principles are a priori. Esa Diaz-Leon, in her critical discussion of AET, thinks this option is plausible for the following reason:

“Maybe if we know, for instance, the individual masses of microphysical entities x₁, x₂ … xₙ, which compose macroscopic entity r, then we can infer a priori the mass of r. This seems plausible because we are using the same predicate both at the microphysical and the macrophysical level, namely, ‘mass’. But what happens when we introduce new predicates at higher-order levels?”

(Diaz-Leon 2011: 106)

Here Diaz-Leon concedes that instances of (AET-1) are a priori given a general principle concerning scale-translation principles. She then attends to the (AET-2) schema, which she finds more problematic (I examine her objection in section 3.3).

Interestingly, Brian McLaughlin appears to reject AET on the basis that mass additivity is logically contingent (I assume McLaughlin agrees that mass additivity is not a plausible example of the contingent a priori, so that if it is contingent it is not a priori):

view that in the fundamental sense, there are no composite entities”. Later I use ‘nihilism’ to refer to the view that there are no composites in any sense.
“Given that the principle of the additivity of mass is logically contingent, the mass of a whole will supervene with only nomological necessity on the masses of its parts; and there are no other properties the parts have taken separately or in other combinations on which the mass of the whole supervenes with logical necessity.”

(McLaughlin 1997: 38)

Unfortunately McLaughlin does not say why he thinks mass additivity is not logically necessary (and hence not a priori). So who is right? I think McLaughlin is right that mass additivity is not an a priori truth. I will give two arguments for the non-apriority of mass additivity. The first argues that our apriority heuristics do not classify mass additivity as a priori. The second argues that the empirical evidence for Einstein's special theory of relativity refutes mass additivity.

The negation of the principle of mass additivity is conceivable. To conceive of mass additivity being false, we simply need to conceive a possibility in which properties of component particles other than mass contribute to the mass of the composite. For example, consider two composites with the same number of parts each having mass m, but one composite is harder to move due to its parts having a higher mean kinetic energy. To conceive this possibility formulate a fictional law that relates the kinematics of particles with their masses and something else (their energies, perhaps) in such a way that their collective behaviour yields this result. Because this possibility is prima facie conceivable, we can see that it is entirely appropriate to empirically test mass additivity. Mass additivity is therefore not a priori.

We can also appeal to modern physics to mount an empirical argument against the claim that mass additivity is a priori. This argument begins with the claim that mass additivity is false and that we empirically discovered it to be false when Einstein discovered the special theory of relativity. According to relativity theory the mass of a composite is related to fundamental properties of its elementary parts by a quite different scale-translation principle than the principle of mass additivity. In particular, composite mass is related to the energies and momenta of component parts in a way that entails that the composites mass is only the sum of its parts in the (rare) case when all of its parts are at rest. I provide the equation for this scale-translation principle in section 2.1. Mass additivity is true in Newtonian worlds. In discovering that our world is not Newtonian but relativistic, we discovered that our world is not a world in which mass is additive. Thus mass additivity is empirically false, while the relativity scale translation principle is empirically true (pending further developments in physics).

31 McLaughlin (1997: 38-9) puts the micro-macro principle of mass additivity on a par with micro-micro principles that state how a particle will interact with other particles, given how it would react with each of those particles individually; his example being the parallelogram law for the composition of forces. I discuss this principle in detail in chapter 2. The non-apriority of such micro-micro principles poses no serious problem for AET* as they can be included in the microphysical description (they don’t mention composites). It is worth noting, however, that the fundamentality of such principles has been questioned: Lange (2009).
These arguments undermine the principle alluded to by Diaz-Leon. The fact that a predicate is applied at both microphysical and macrophysical scales implies nothing about the epistemic status of principles that relate their application at those scales. They also help to undermine the first step in Chalmers and Jackson’s argument for AET*, as we have seen. So, in the absence of any obvious way to a priori infer the mass of composites from microphysics, should we reject AET* and hence AET? I think not. For although we have seen that the two options considered by Chalmers and Jackson are not viable, there is a third option: a posteriori scale-translation principles themselves follow a priori from microphysical descriptions.

Although McLaughlin is right that mass additivity is not a priori, the claim that there are no other properties of the parts taken together or in other combinations, which a priori determines the mass of the whole, is wrong. I will argue that the mass of the whole is determined a priori by the following properties of the parts: having certain masses, forces, kinematics and obeying the fundamental laws of nature. I motivate this idea by appeal to a simple Newtonian possible world. I argue that mass additivity follows a priori from its fundamental microphysical description. The a priori material conditional I argue for is:

\[(\text{Mass additivity}): \text{If } [\text{Newtonian microphysical description}] \& [T] \text{ then mass is additive.}\]

This is a highly non-trivial conditional and consequently its epistemic status is difficult to determine. And so we are still left with a methodological problem: the goal is to defend AET from the bottom up, by arguing for the apriority of (Mass Additivity). But without the comfort of a priori scale-translation principles, it is not at all obvious where we should begin. Applying our apriority heuristics from section 1.1, to (Mass Additivity), is going to be unwieldy. For example, before we even ask whether its negation is conceivable, we will need to flesh out the details of the antecedent. What we need is a way to break down complex conditionals such as (Mass additivity) so that they are easier to evaluate. In the next section I argue that \textit{a priori entailment expansions} can achieve this.

\section*{1.5 A Priori Entailment Expansions}

To illustrate the idea of an a priori entailment expansion, consider the following conditional:

\[(\text{Dots-to-shapes}): \text{If } [\text{dot-distribution description}] \text{ then there is a square}^{32} \]

\[^{32}\text{Pettit (1994) considers a similar case.}\]
Consider an instance of the antecedent of (Dots-to-shapes) such that its complexity makes it difficult to determine whether it a priori entails a square. The antecedent might nonetheless be simple enough to allow us to see that it a priori entails certain \textit{lines}. In that case, we first a priori infer the lines, and then see if the lines a priori entail a square. We can model this in the form of a deductive argument:

\begin{align*}
(1) & \quad \text{[dot-distribution description]} \\
(2) & \quad \text{If (1) then [lines description]} \\
(3) & \quad \text{If (1) \& (2) then there is a square} \\
(4) & \quad \text{Therefore, there is a square}
\end{align*}

The antecedent of (Dots-to-shapes) is the first premise of the deductive argument while the consequent of (Dots-to-shapes) is the argument’s conclusion. Premises (2) and (3) are instances of what I call \textit{expansion conditionals}. If one can show that the expansion conditionals are a priori then one has thereby shown that (Dots-to-shapes) is a priori. Expanding a conditional out into a deductive argument in this manner yields an \textit{a priori entailment expansion}. A priori entailment expansions enable us to argue that complex conditionals are a priori by arguing that a number of simpler expansion conditionals are a priori.

Let’s illustrate this using a concrete expansion of (Dots-to-shapes):

\begin{itemize}
  \item[(1)] There is a two-dimensional flat surface with small (relative to the surface) black dots distributed over points on that surface. Using Cartesian coordinates, there is a continuous distribution of dots going from $(x_{10}, y_{10})$ to $(x_{20}, y_{10})$.\textsuperscript{33} There is another going from $(x_{20}, y_{10})$ to $(x_{20}, y_{20})$. There is another going from $(x_{20}, y_{20})$ to $(x_{10}, y_{20})$. And there is another going from $(x_{10}, y_{20})$ to $(x_{10}, y_{10})$. Other than the surface, these dots, and whatever (if anything) they compose, there is nothing else.

For clarity I will break up premise (2) from above into four distinct premises, where each premise expresses an expansion conditional that infers a particular side of the square:

\begin{itemize}
  \item[(2)] If (1) then there is a straight line (composed of dots) going from $(x_{10}, y_{10})$ to $(x_{20}, y_{10})$.
  \item[(3)] If (1) then there is a straight line (composed of dots) going from $(x_{20}, y_{10})$ to $(x_{20}, y_{20})$.
  \item[(4)] If (1) then there is a straight line (composed of dots) going from $(x_{20}, y_{20})$ to $(x_{10}, y_{20})$.
  \item[(5)] If (1) then there is a straight line (composed of dots) going from $(x_{10}, y_{20})$ to $(x_{10}, y_{10})$.
\end{itemize}

Premises (2) to (5) are expansion conditionals that are almost certainly a priori, so we have a priori inferred four lines from (1). Now, we can just think about what shape, if any, these lines a priori entail.

\textsuperscript{33} Think of ‘there is a continuous distribution of dots going from $(x_{10}, y_{10})$ to $(x_{20}, y_{10})$’ as being shorthand for ‘there is a dot at $(x_{10}, y_{10})$, another at $(x_{11}, y_{10})$, and another at $(x_{12}, y_{10})$, \ldots, and another at $(x_{20}, y_{10})$.}
entail. And here, we simply need to reflect on where the ends of the lines all meet to infer that there is a probably a square.

But there might not be a square! For what if our surface contains not only these lines, but a continuous infinity of other lines smeared all over these four lines? In that case we might reasonably judge that there is no square. Here, we must appeal to the “that’s all” clause in premise one. In particular, after inferring that the four lines probably compose a square, we must go back to premise one and check for justification defeaters. The “that’s all” clause a priori entails that there are no such defeaters, because there are no other entities (dots) that could compose such lines.

Our final expansion conditional, which gets us to our desired conclusion, references premises (1) to (5):

(6) If (1)-(5) then there is a square.
(7) Therefore, there is a square.

The goal was to show that the conditional ‘If [dot-distribution description] then there is a square’ is a priori. The complexity of the antecedent made this goal difficult. To resolve this we expanded the conditional out into a deductive argument, where the consequent became the conclusion, the antecedent became the first premise, and a number of expansion conditionals were added as extra premises to ensure the validity of the argument. The expansion conditionals represent inferences that one might draw from the first premise, so as to infer the conclusion in a step by step manner. Their apriority is relatively transparent and if they are a priori then the expanded conditional is too. This is how one constructs an a priori entailment expansion.

There is an additional strategy we can employ to simplify the task of evaluating the epistemic status of complex conditionals. Imagine there is dispute over whether ‘If X then Y’ is a priori (for some X and Y). Before setting out an a priori entailment expansion, we might be able to simplify the antecedent in a way that makes it easier to resolve the dispute. In that case, to resolve dispute over whether ‘If X then Y’ is a priori we can expand the conditional ‘If X* then Y’ where X* is a simplified version of X. The basic idea is that if we are justified in asserting that ‘If X* then Y’ is a priori and X* and X bear no relevant difference then we are justified in asserting that ‘If X then Y is a priori. To ‘bear no relevant difference’ in this context means that the only difference between X and X* is the sort of complexity that makes no difference to a priori entailment.

Frank Jackson (1998) appeals to a similar idea. Jackson argues that the following conditional is a priori: ‘If [physical description of human organism O] then O has psychological properties’. Jackson’s argument appeals to the premise that the following simpler conditional is a priori: ‘If [physical description of amoeba A] then A has such and such properties’. His argument is worth quoting in full:
“[T]here is a much shorter way of making plausible the idea that physicalism is committed to the a priori deducibility of psychological nature from physical nature. It is implausible that there are facts about very simple organisms that cannot be deduced a priori from enough information about their physical nature and how they interact with their environments, physically described. The physical story about amoebae and their interactions with their environment is the whole story about amoebae. [...] Now, according to physicalism, we differ from amoeba essentially only in complexity of ingredients and their arrangement. It is hard to see how that kind of difference could generate important facts about us that in principle defy our powers of deduction. Think of the charts in biology classrooms showing the evolutionary progression from single-celled organisms on the far left to the higher apes and humans on the far right: where in that progression can the physicalist plausibly claim that failure of a priori deducibility of important facts about these organisms and creatures emerges?”

(Jackson 1998: 83-84)

The problem with Jackson’s argument is that the initial premise is very controversial: it is not implausible that facts about very simple organisms cannot be deduced a priori from enough information about their physical nature and how they interact with their environments, physically described. I don’t intend to try to show that all facts about amoebae follow apriori from physical descriptions (a difficult task). But I do intend to do something analogous, if less ambitious. I show that salient Newtonian macrophysical properties follow a priori from simple Newtonian worlds. However, amoebae differ from the macrophysical properties I hope to derive only in complexity. So if my argument works, then we have an argument for Jackson's initial assumption and we can therefore employ his reasoning more generally.

In chapters two and three I will expand complex conditionals in accordance with the following algorithm for constructing a priori entailment expansions:

**Step One:** Simplify the antecedent down as much as possible—removing information that is tangential to the dispute over its epistemic status.

**Step Two:** Set out an argument that has the consequent as its conclusion and the (simplified) antecedent as its initial premise.

**Step Three:** Insert a set of further premises that (a) are themselves material conditionals, which model a set of basic a priori inferences that one might draw if one were to try to infer the conclusion from premise one, (b) have relatively transparent epistemic statuses, and (c) validate the argument. Call these expansion conditionals.
**Step Four**: Argue that the expanded conditional is a priori by arguing that the expansion conditionals are a priori.

The complex conditionals I expand in chapter two are instances of (Mass additivity). The goal will be to expand these conditionals to show that Newtonian worlds provide some level of confirmation of AET.
A Priori Entailment as a Constraint on Classical Physical Theories

2.1 Mass Additivity

The principle of mass additivity states that the mass of a composite object is the sum of the masses of its atomic component parts. In this section I provide five reasons why mass additivity is an important test case for AET, before summarising what is to come.

Firstly, there are real attempts to reductively explain mass additivity in terms of fundamental microphysical truths of Newtonian worlds (section 2.3). If these attempts are successful then physicists have shown that mass additivity is grounded in Newtonian microphysics. It is then clear how to use mass additivity as a test case for AET: determine whether the relevant Newtonian microphysics a priori entails mass additivity (e.g. with an a priori entailment expansion).

Secondly, discussion of mass additivity in current AET literature is problematic. Advocates and critics have treated mass additivity as an a priori truth, but it is a posteriori false (section 1.3). Chalmers and Jackson (2001: 331) suggest that microphysical truths a priori entail macrophysical truths in virtue of the apriority of scale-translation principles such as mass additivity. In her critique of AET, Diaz-Leon (2011: 106) agreed because the principle of mass additivity relates the application of the same predicate at both micro and macro scales. But mass additivity is a posteriori false. According to special relativity, a more complicated principle relates the mass of a composite to microphysical properties. Where \( m_c \) is the mass of the composite composed of two particles, with masses \( m_1, m_2 \) and momenta \( p_1, p_2 \), respectively, the scale-translation principle is:

\[
m_c = \left( \sqrt{m_1^2 + \frac{p_1^2}{c^2}} + \sqrt{m_2^2 + \frac{p_2^2}{c^2}} \right)^2 \left( \frac{p_1 + p_2}{c^2} \right)^2 - \frac{1}{2}
\]

Equation (1) can also be written more generally, where \( m_c \) is the mass of the composite composed of \( n \) particles, each with energy \( E_i \) and momenta \( p_i \),

\[
m_c = \left( \sum_{i=1}^{n} \frac{E_i}{c^2} \right)^2 - \left( \sum_{i=1}^{n} \frac{p_i}{c} \right)^2 \right)^{1/2}
\]

\[
\text{(2)}
\]

34 Khrapko (2000).
In discovering that our world is not Newtonian but relativistic, we discovered that our world is not a world in which mass is additive. Thus mass additivity is a posteriori false, while the relativity scale translation principle is a posteriori true (pending further developments in physics). It is therefore important to determine how AET should account for mass additivity in worlds in which it is true, such as Newtonian worlds. I discuss worlds in which mass additivity is false in section 2.8.

Thirdly, some refer to what I call ‘mass’ as ‘invariant mass’ (or ‘rest mass’ or ‘proper mass’) to distinguish it from what they call ‘relativistic mass’. Relativistic mass is additive whereas invariant mass obeys the above equations. I join most physicists in rejecting the notion of relativistic mass and consider invariant mass (i.e. mass) to be the referent of Newton’s term ‘mass’.36 I provide a novel defence of this view in section 2.2. I argue that the concept of mass is constitutively connected to the concepts of position, motion, and force. For this reason, the case of mass additivity allows for a comprehensive treatment of AET in relation to Newtonian worlds. I argue furthermore that the concept of mass is associated with a naturalness constraint, and that this helps to explain why we now think mass additivity is false.

Fourthly, because the case of mass additivity requires analysis of a host of related concepts, we can generalise from it. This is useful when analysing other reductions e.g. in relativity theory or quantum mechanics. The mass additivity example works as a paradigmatic exemplar. Once we have in place an expansion that demonstrates the a priori entailment of mass additivity and other macrophysical properties, we will have an illustrative example showing how one can demonstrate similar a priori entailments for a variety of physical theories.

Fifthly, and finally, the case of mass additivity is of interest in itself and I believe that the a priori entailment expansion(s) I provide for (Mass additivity) constitute the most comprehensive explanation of mass additivity in Newtonian mechanics that has ever been proposed. In particular, I argue that the most extensive explanation currently in print (Feather (1965)) begs the question by assuming that the position of a composite is identical to its centre of mass (section 2.3), and I provide an explanation that avoids that assumption (section 2.4) and extend the explanation to gravitational mass (section 2.5).

The primary goal of this chapter is to defend the apriority of instances of the (Mass additivity) schema:

(Mass additivity): If [Newtonian microphysical description] & [T] then mass is additive.

36 For example, Einstein (letter reproduced in Okun (1989)), Taylor and Wheeler (1966), Gabovich and Gabovich (2007), and Okun (2009a). Oas (manuscript) documents the historical decline in the use of ‘relativistic mass’ in physics textbooks and papers. For a contrary view see Field (1973), who argues that Newton’s term ‘mass’ has indeterminate extension due to the equal legitimacy of ‘invariant mass’ and ‘relativistic mass’. Sandin (1991) defends ‘relativistic mass’ through its usefulness in teaching relativity to beginners. For an overview of the debate see Jammer (2000, chapter 2).
There are two primary concepts of mass in Newtonian mechanics: gravitational mass and inertial mass. Inertial mass is roughly defined as the property that disposes systems to resist changes in motion given applied forces. Meanwhile, gravitational mass is roughly defined as the property of objects that causes them to exert an attractive force on other (gravitationally massive) objects. Gravitational mass and inertial mass are conceptually distinct, but a posteriori equivalent. If this equivalence principle is fundamental in Newtonian worlds then we can infer the additivity of inertial mass from Newtonian microphysics and then infer the additivity of gravitational mass simply by adding the equivalence principle to the microphysical truths. I will be more thorough and provide expansions for both additivity principles individually.

This yields two distinct (Mass additivity) schemas:

(Inertial mass additivity): If [Newtonian microphysical description] & [T] then inertial mass is additive.

(Gravitational mass additivity): If [Newtonian microphysical description] & [T] then gravitational mass is additive.

I also defend the apriority of instances of (Charge additivity):

(Charge additivity): If [Classical microphysical description] & [T] then charge is additive.

This yields a relatively comprehensive defence of AET in classical physical worlds (i.e. worlds described by pre-relativistic and pre-quantum Newtonian point-particle mechanics). I also sketch how the following conditional can be expanded:

(Invariant mass non-additivity): If [relativistic microphysical description] & [T] then invariant mass is non-additive.

### 2.2 The Concept of Mass

In this section I argue that the concept of mass is inferentially connected to other physical concepts such as position, motion, and force, and to the metaphysical concept of naturalness. Here, a concept is inferentially connected to the concept of naturalness if and only if the concept is applied (where possible) to properties resembling fundamental properties. I argue that an implicit appreciation of these inferential connections is a requirement for fully possessing the concept of mass. I argue that these inferential connections explain why we now reject mass additivity. These arguments enable me in later sections to argue that sufficient information about the positions, motions, and forces of objects
enables us to a priori infer their masses. I also distinguish the concept of inertial mass from gravitational mass.

In his paper *On Defining Mass*, Eugene Hecht considers the standard definition of ‘mass’:

“A sophisticated definition appearing in countless textbooks and classrooms is based on the idea of inertia. The mass of an object is a measure of, and gives rise to, its resistance to changes in motion; \( F = ma \), which stands on a rich experimental history, presumably quantifies the traditional idea of “inertia”.

(Hecht 2011: 40)

The standard definition of ‘mass’ in terms of resistance to changes in motion encapsulates the basic idea that the harder an object is to push around in space, the more massive it is. Once we realise that objects have some intrinsic property responsible for this resistance to being pushed around, we introduce a word—‘mass’—to apply to that property. If this is correct then we can test whether a subject understands the word ‘mass’, or whether a subject possesses the concept of mass, by determining whether the subject has a basic appreciation of the connection between mass and resistance to changes in motion.

We must distinguish *inertial mass* from *gravitational mass*. Prior to possessing the concept of gravitational mass, we possessed the concept of *weight*. Weight is the property of objects that causes us to distinguish light objects from heavy objects, according to the muscular effort required to lift or support them. Our concept of weight is a pre-theoretical concept designating the aspect of gravitational mass that can be directly perceived. Our concept of gravitational mass is theoretical: we have to theorise about the underlying nature of weight to put ourselves in a position to form the concept of gravitational mass.

Like weight, the concept of inertial mass is also pre-theoretical. Pockman (1951: 305) notes that “The use and popular connotation of our words “heft” or “hefty” in contrast to the connotation of “weight” or “heavy,” is a layman’s reflection of this intuitive idea of inertia.” Pockman then discusses an illuminating etymological finding in Dampier (1940: 37):

“Man’s sense of inertial mass and momentum became much less vague, more quantitative, with the development of the potter’s wheel, the bow and arrow, the catapult, and finally, by Galileo’s time, firearms. In this connection, it is of interest to note that according to Dampier, Balliani, captain of archers at Genoa, was the first man to make a distinction between mass and weight; he distinguished *moles* from *pondus*. Balliani was a contemporary of Galileo. Dampier also notes that Leonardo, a century earlier, had said, “Every body has a weight in the direction of its movement.””
Just as the concept of inertial mass becomes clear when one realises that objects have some property responsible for their varying dispositions to resist changes in motion, so too does the concept of gravitational mass become clear when one realises that objects have some property responsible for their varying gravitational attractions (other than relative position). One weighs less on the moon in virtue of the moon having a weaker gravitational pull than Earth. The moon has a weaker gravitational pull in virtue of having less gravitational mass. Thus, the gravitational mass of an object is the property responsible for the object’s gravitational strength—its strength to attract (by accelerating) other objects in its vicinity.38

Gravitational mass and inertial mass are clearly distinct concepts. To see this, notice that one can conceive an object which has no gravitational pull on other objects, but is incredibly difficult to set into motion and stubbornly resists changes in its motion given applied forces. We know only a posteriori that gravitational mass and inertial mass are identical. We have good pre-theoretical reason to believe this: our interactions with objects suggest that the extent to which objects are difficult to set in motion is proportional to how difficult they are to lift. Newton gave a scientific reason. Newton discovered empirically that inertial mass \( m \) obeys the equation \( a = F/m \) and that gravitational mass \( m' \) obeys the equation \( F = Gm_1m_2/r^2 \). If we assume that \( m'_1 = m \) then we may derive \( a = Gm'_2/r^2 \) which quantifies what we know by experiment: that the acceleration of a freely falling body is independent of its mass \( m'_1 \) and depends only on (i) the mass of the body accelerating it \( (m'_2) \), (ii) its distance from that body \( (r) \) and (iii) the gravitational constant G.

The inertial mass definition is controversial and Hecht (quoted above) only introduces it to criticise it. Hecht argues that in relativity theory, the idea of inertia cannot be precisely quantified into an equation such as Newton’s second law (force equals mass times acceleration):

\[
F = ma
\]  

(3)

In particular, \( F=ma \) becomes a low speed approximation, and the mass of an object is no longer a proportionality constant between \( F \) and \( a \). An object's resistance to changes in motion given a force does not simply depend upon its mass, but also upon the direction of the force relative to the direction of the object's velocity. For example, when the velocity and the direction of force are perpendicular, we have:

37 The distinction may have been made explicit much earlier. As Hecht (2006: 41) writes: “In the 13th century the theologian Aegidius Romanus, while considering the Eucharist, suggested that in addition to weight and volume there was a third measure of matter, the quantitas materiae or "quantity of matter."”

38 Jamer (2000: 6) notes that we can distinguish two concepts of gravitational mass; active gravitational mass (a measure of the strength of the gravitational field produced by a body) and passive gravitational mass (a measure of a body’s susceptibility to a given gravitational field).
\[ F = \gamma ma \] (4)

And when the velocity and the direction of force are parallel, we have:

\[ F = \gamma^3 ma \] (5)

Where:

\[ \gamma = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}} \] (6)

Here \( v \) is the particle's velocity, \( c \) is the speed of light and \( \gamma \) is the gamma factor (or Lorentz factor). The reason for this peculiar phenomenon is that in relativity theory, objects cannot accelerate past the speed of light. So in relativity theory, an object’s disposition to resist changes in motion given an applied force is not merely determined by the mass of the object, but also by its velocity. And furthermore, due to the geometry of spacetime, which prevents objects travelling faster than light, the direction of velocity relative to the applied force is also a determining factor.

Hecht concludes from this that “mass is not identically inertia” and adds that “that doesn’t mean that mass and inertia are not connected concepts; they surely are. It just means there is no simple relationship that informs a straightforward definition of mass” (2011: 40). Hecht goes on to define mass “from a contemporary perspective” in terms of relativistic notions that are not applicable to Newtonian worlds: “The invariant mass of any object—elementary or composite—is a measure of the minimum amount of energy required to create that object, at rest, as it exists at that moment” (2011: 43).

The word ‘definition’ is used in a variety of ways and different theorists employ different constraints for its application. Hecht admits that mass and inertia are conceptually connected but does not use the notion of inertia in his definition of mass because inertia is not precisely quantifiable into a simple relativistic equation. This is a strong constraint to put on a definition and it is not surprising that the definition of ‘mass’ that Hecht ultimately provides has little relation to the use of ‘mass’ in Newtonian physics, and pre-theoretical discourse. Our goal is to capture the aspect of the term’s meaning that stays fixed throughout significant empirical discoveries, so this won’t do.

If a definition captures the aspect of meaning that stays fixed throughout significant empirical discoveries, then the definition can help explain the growth of knowledge. For consider the following sentence: ‘we used to think that mass was an additive property but now we know otherwise’. This sentence only makes sense if ‘mass’ means the same in Newton's theory as in Einstein's theory. If the meaning of ‘mass’ in Newtonian mechanics differs from the meaning of ‘mass’ in relativity theory, then relativity theory has not advanced our understanding of mass, but rather, has changed the topic.
But this is surely not the case. After all, if Leonardo da Vinci was transported from his time (1452-1519) into our time, we would like to be able to tell him interesting things about that property he referred to as “weight in the direction of motion”, such as its non-additivity. So our definition of ‘mass’ should retain reference to inertia. But something needs to be added to make sense of why mass and inertia come apart in relativity theory. Here, the notion of *naturalness* plays an essential role. Applying terms to the most natural properties possible means applying terms to the properties that are most like fundamental properties.

In the history of relativity theory, the quantity $\gamma m$ in (4) was referred to as a kind of mass: *longitudinal* mass $m_L$; and the quantity $\gamma^3 m$ in (5) was also referred to as a kind of mass: *transverse* mass $m_T$. More formally, the relativistic expression relating force and acceleration for a particle with non-zero mass $m$ moving in direction $x$ is:

\[
F_x = \gamma^3 m a_x = m_L a_x \tag{7}
\]

\[
F_y = \gamma m a_y = m_T a_y \tag{8}
\]

\[
F_z = \gamma m a_z = m_T a_z \tag{9}
\]

It is clear why the notions of transverse and longitudinal mass were introduced: mass is conceptually connected to inertia, and transverse and longitudinal mass exactly quantify the inertia of a particle. So why introduce the notion of mass $m$ that has no subscripts, in (7), (8), and (9), if it does not exactly quantify inertia? Why not call it something else? The answer is that inertia is not the only important property of our concept of mass that determines its reference. We also use ‘mass’ to refer to the most *natural* property responsible for inertia. Transverse and longitudinal mass are non-natural properties, which are direction dependent and velocity dependent. Their exact values depend (arbitrarily) on frame of reference and are therefore not objective properties as fundamental properties should be.

At least two features of relativity force us to move away from equations (7)-(9). Firstly, mass is not constant in relativity theory: a collision can alter the mass of a particle. Secondly, the force terms in equations (7)-(9) are three vectors specified in terms of three numbers corresponding to the magnitude of the force along each of three spatial directions. But the fundamental geometry of relativity theory is four-dimensional, requiring the use of four-vectors.

Let’s consider the first feature in more detail. In Newtonian mechanics, the second law (equation (3)) is equivalent to an expression equating the force to the rate of change of momentum $p$:

\[
F = \frac{dp}{dt} = \frac{d(mv)}{dt} \tag{10}
\]
Equations (3) and (10) are equivalent in part because the $m$ in $d(mv)$ is constant (does not change over time). Because $m$ is constant it can be pulled out of the time derivative leaving:

$$ F = m \frac{d\dot{v}}{dt} \quad (11) $$

And the time derivative of velocity is acceleration, by definition. In relativity, one cannot move from (10) to (11) because mass is not constant.\(^3^9\) One instead stays with the relativistic (four-vector) version of (10). Because mass is not constant given applied forces, one should not look to the proportionality constant between force and acceleration to find what plays the inertial mass role for particles, one should instead look to the proportionality constant between momentum and velocity.

In relativity the proportionality constant is the product of “$m$” and the gamma factor:

$$ p = \gamma mv \quad (12) $$

Force is equivalent to the time derivative of the product of “$m$” and $\gamma$:

$$ F = \frac{dp}{dt} = \frac{d(\gamma mv)}{dt} \quad (13) $$

So $\gamma m$ plays the inertial role in relativity. Accordingly, $\gamma m$ has a history of being referred to as relativistic mass. Indeed, many 20th century textbooks distinguish $m_R$ from $m_0$ where $m_R$ is the relativistic mass, and $m_0$ is rest mass: the relativistic mass $m_R$ that the body would have if its velocity were 0 ($\gamma = 1$).\(^4^0\)

The velocity of an object must be defined by appeal to some arbitrary frame of reference. And so because relativistic mass ($m_R = mv = m_0 \nu$) depends on velocity, the relativistic mass of an object depends on some arbitrary frame of reference. Properties that depend on arbitrary human impositions such as reference frames are not natural properties. And so insofar as we are using ‘mass’ to refer to the most natural property responsible for an object’s inertia, we should find a better referent than the relativistic mass $m_R$. And this is precisely what has happened in modern physics. The distinction

\(^3^9\)“But they are equivalent only if the rest mass is constant, which is by no means always the case. For example, if two particles collide elastically, their rest masses during collision will vary, but that would be precisely when we might be interested in the elastic forces acting on them” (Rindler 1991: 102).

\(^4^0\)Oas (manuscript) reviews 637 works of physics and determines that 477 use the notion of ‘relativistic mass’. On historical analysis Oas concludes that the use of ‘relativistic mass’ has been diminishing in introductory textbooks, has stayed constant in more advanced relativity texts, and is growing in more informal works (e.g. popularizations of physics). Oas excluded peer-reviewed journal articles from the survey to simplify the task and to limit to sources accessible to non-scientists. Lev Okun, who himself has published over 300 articles, asserts that “As regards active specialists they answer in perfect unison insofar as their scientific work is concerned: the mass is independent of velocity, it is not additive […] there is no disagreement among researchers on the definition of mass. […] According to modern terminology, both terms, ‘relativistic mass’ and ‘rest mass’, are obsolete. (Okun 2000: 1270).
between relativistic mass $m_R$ and rest mass $m_0$ has been largely abandoned in favour of a single mass $m$ which refers to the frame invariant property labelled $m$ in equations (12) and (13).

We are now in a better position to understand why mass is not additive in relativity theory. Relativistic mass $m_R$ is additive in relativity theory. But relativistic mass is not a natural candidate for being the referent of ‘mass’. The most natural referent is whatever frame-invariant property contributes the most to an object’s inertia. This property is the measure of the inertia of a body when the body is at rest ($\gamma = 1$). Now, let’s say we distribute some values to some particles, to quantify their inertial dispositions in their rest frames. The sum of these values is not in general going to be a measure of their composite’s inertial disposition in its rest frame. That’s because in most cases, once we’ve determined the composite rest frame, we find the components moving with some velocity relative to the rest frame. And the composite’s inertial disposition in its rest frame depends on its parts movements relative to that frame. This is experimentally confirmed: take a box of gas, heat it up so that the gas particles move faster relative to each other, and the box offers more inertia, or is more massive (Taylor and Wheeler 1966). Mass is only additive when component particles are at rest. For particles in motion, mass is slightly non-additive if they are moving slowly relative to each other, but drastically non-additive otherwise (Okun 2009b: 431).

Thus, to understand why we now believe that mass is non-additive, it is crucial that we interpret our use of mass in terms of a naturalness constraint. By definition, inertial mass is that property of bodies that is the most natural contributor to their dispositions to resist changes in motion given applied forces. I will say in various places throughout the dissertation that if a property is the most natural measure of the disposition for motion resistance, then that property plays the inertial mass role. Possessing a concept means that one is in a position to pick out what objects in the environment play the conceptual role for that concept, given sufficient information about the environment. Thus, given sufficient information about how an object resists motion given applied forces, one will be able to deduce the inertial mass of the object.

The concept of gravitational mass should be given a similar treatment. By definition, gravitational mass of an object is a measure of its most natural contributor of gravitational attractions on objects as they come closer. I will say more about gravitational mass in section 2.5. For now, let us return to the classical context, and consider how physicists try to reductively explain why mass is additive in Newtonian worlds. This will act as a guide to the metaphysical grounds of mass additivity in Newtonian worlds. We can then ask whether these grounds a priori entail mass additivity as a test case for AET.
2.3 The Reductive Explanation of Mass Additivity

Mass additivity is often presupposed, rather than explained, in Newtonian physics textbooks. This is surprising given the important roles that composite mass plays. For example, the motion of a composite is typically calculated, by calculating the motion of the composite’s centre of mass. This is possible because there is a proof demonstrating that the composite’s centre of mass obeys Newton’s laws. As we shall see at the end of the section, this proof presupposes mass additivity. Furthermore, a microphysical explanation of mass additivity is necessary if one is to fully understand Newton’s theory. A similar point is made by Norman Feather:

“It is not customary, in undergraduate texts, to distinguish more than formally between the Newtonian mechanics of particles and the mechanics of gross bodies. But Newton was an atomist, and it is clear that he conceived of the forces which he identified as operating between bodies, and whose interrelations he described in the third law, as compounded of the elementary forces acting between the particles (primordial atoms) of which real bodies are constituted. [...] It is clear, then, that the question of the additivity of mass is fundamental for the logical development of the Newtonian scheme: without its consideration, the transition from particles to gross bodies cannot be made convincingly, and the simple mechanics of particles fails to achieve relevance to the real world.”

(Feather 1965: p511)

Some textbooks do provide reductive explanations of Newtonian inertial mass additivity and the explanations typically take the same general form. I have never come across an explanation of gravitational mass additivity. I will say why I believe this to be the case in a moment as it will prove to be important. Until then, let every instance of ‘mass’ in what follows refer to inertial mass.

The following explanation, taken from Kibble and Berkshire’s *Classical Mechanics*, aims to show that “one consequence of our basic laws is the additive nature of mass” (2004: 12). It is clear that Kibble and Berkshire are offering a reductive explanation that answers a grounding question. After all, their notion of ‘consequence’ is not a causal notion—they are not trying to show that a composite’s mass is a causal consequence of anything. Rather, they aim to show that mass additivity is grounded in, or obtains in virtue of, the basic laws. Hence, they propose to reduce mass additivity to the basic laws.

Kibble and Berkshire consider an isolated three body system in light of Newton’s second law ($F_i = m_i a_i$) and the superposition principle for interactions ($F_i = \sum_{j=1}^{N} F_{ij}$). The former law states that the force on object $i$ is equivalent to the product of the object’s mass and acceleration. The latter
law states that the force on object \( i \) is equivalent to the sum of each of the individual forces \( F_{ij} \), where \( F_{ij} \) stands for the force on object \( i \) due to the object indexed by \( j \) \((j=1,2,...,N)\). Applying these two laws to each of the three particles gives:

\[
F_{12} + F_{13} = m_1 a_1 \tag{14}
\]

\[
F_{21} + F_{23} = m_2 a_2 \tag{15}
\]

\[
F_{31} + F_{32} = m_3 a_3 \tag{16}
\]

Now the third law \((F_{ij} = -F_{ji})\), which states that there is a force on object \( i \) due to object \( j \) if and only if there is an equal and opposite force on \( j \) due to \( i \), is introduced. The third law is introduced to show that if we add these equations then the terms on the left cancel in pairs, such that:

\[
m_1 a_1 + m_2 a_2 + m_3 a_3 = 0 \tag{17}
\]

A simple algebraic transformation then gives an equation for \( m_1 a_1 \):

\[
m_1 a_1 = -m_2 a_2 - m_3 a_3 \tag{18}
\]

Now comes a crucial simplification:

“if we suppose that the force between the second and third is such that they are rigidly bound together to form a composite body, their accelerations must be equal: \( a_2 = a_3 \). In that case, we get

\[
m_1 a_1 = -(m_2 + m_3) a_2 \tag{19}
\]

which shows that the mass of the composite body is just \( m_{23} = m_2 + m_3 \)” (2004: 12).

Why do Kibble and Berkshire think that equation (19) shows that the composite’s mass is additive? After all, (19) merely relates properties of the composite’s atomic parts. Why should the fact that an algebraic transformation that contains the expression ‘\((m_2 + m_3)\)’ explain anything at all about the composite of particles 2 and 3?

The basic idea must be this: when a composite’s parts each have acceleration \( a \) then the composite has acceleration \( a \). So the right hand side of equation (19) contains a term for the composite’s acceleration yielding: \( m_1 a_1 = -(m_2 + m_3) a_c \). Now consider the left hand side of (19). If we could derive \( m_1 a_1 = -F_{C1} \) then we would have a crucially important expression relating the force on the composite and the composite acceleration to the sum of the masses of its parts. In particular we could derive:
Thus, granting the assumption that $m_1a_1 = -F_{C1}$, Kibble and Berkshire have essentially worked out the force and acceleration of the composite and have then used the second law to solve for its mass, the value of which is suggested by equation (19), and is made clear by equation (20).

I will return to this explanation in the context of evaluating RET in section 3.1. For now, I simply use the explanation as a guide for determining what the permissible additions to our entailment base are, if we are to use mass additivity as a test case for AET. The explanans explicitly appeals to four key resources, all of which have a claim to being fundamental aspects of Newtonian worlds:

**Newton's second law:** $F_i = m_ia_i$

**Newton's third law:** $F_{ij} = -F_{ji}$

**The superposition principle:** $F_i = \sum_{j=1}^{N} F_{ij}$

**Simple physical states of point particles:** The accelerations and masses of closed systems of point particles.

Kibble and Berkshire's reductive explanation does not exhibit a priori entailment. That is, the explanans does not a priori entail the explanandum. This is partly because the explanans does not a priori entail the masses of composites whose parts do not have identical accelerations. Furthermore it is not yet clear how to derive the force exerted by the composite or the force exerted on the composite. This poses a clear problem for RET, or for any theory of reductive explanation that requires reductive explanations to exhibit a priori entailment. This does not pose an obvious problem for AET, however. AET only requires fundamental Newtonian descriptions to a priori entail mass additivity. The fact that the textbook explanation does not immediately demonstrate how this is possible does not show that it is impossible.

There is also a problem with Kibble and Berkshire’s explanation as it stands: it is not clear what the analogue explanation of gravitational mass additivity is supposed to be. This is important because according to Kibble and Berkshire, mass additivity is grounded in Newton’s fundamental laws. But it would be strange if inertial mass additivity is explainable in terms of Newton’s laws but gravitational mass additivity is not.

---

42 To infer $m_1a_1 = -F_{C1}$, one option would be to start by inferring that C exerts F on particle 1 in virtue of C’s components exerting F on particle 1. The third law applied to C and particle 1 would then entail that the force on C is $-F_{C1}$. However, it is not clear that the third law can be applied to more than two particles like this. In their discussion of the third law Kibble and Berkshire say, “It would be perfectly possible to include also, say, three-body forces, which depend on the positions and velocities of the three particles simultaneously. However, within the realm of validity of classical mechanics, no such forces are known, and their inclusion would be an inessential complication” (2004: 7). In that case, we should ask whether there is some other way to derive $m_1a_1 = -F_{C1}$. I provide a derivation in the next section.
Kibble and Berkshire required the simplification that \( a_2 = a_3 \). The analogous simplification in the gravitation case is position identity among components. To see this, consider Newton's gravitation law. The magnitude of the gravitational field \( g(x) \) that a single point mass \( (m_1) \) located at \( x_1 \) determines for some position \( x \) is:

\[
g(x) = G \frac{m_1}{|x - x_1|^2} \tag{21}
\]

Here, \( G \) is the gravitational constant. If we know \( g(x) \) then we can calculate the magnitude of the force that a test particle with a certain mass would feel if it were located at \( x \). The straight brackets remove vectorial information regarding the field value, retaining only information about the magnitude. To account for this missing information we express the law as follows:

\[
g(x) = G \frac{m_1(x - x_1)}{|x - x_1|^3} \tag{22}
\]

The law for \( N \) masses is analogous to the superposition of forces law. It is a composition law for field contributions, stating that the total value of the gravitational field at a position \( x \) is the sum of the values determined by each individual mass (indexed by \( i \)):

**The Composition of Fields:** \( g(x) = \sum_{i=1}^{N} g(x)_i \)

With this in mind, the gravitation law for \( N \) masses is:

\[
g(x) = G \sum_{i=1}^{N} \left[ \frac{m_i(x - x_i)}{|x - x_i|^3} \right] \tag{23}
\]

This states that the value of the gravitational field at point \( x \) as determined by multiple masses \( (m_i) \) is the sum of each individual value that would be determined by each mass individually in accordance with equation (22).

Let’s say we have two particles (so that \( N = 2 \)). In that case, equation (23) becomes:

\[
g(x) = G \frac{m_1(x - x_1)}{|x - x_1|^3} + G \frac{m_2(x - x_2)}{|x - x_2|^3} \tag{24}
\]

Unlike the equation for Newton’s second law, the gravitation equation is not a function from masses and accelerations to forces. Rather, it is a function from mass and position to force (or to force fields). So the analogue of treating particles as having identical accelerations is to treat particles as having identical positions. So if we assume that \( x_1 = x_2 \), then the gravitation analogue of equation (19) is:

---

43 I use this equation, rather than the equation for the gravitational force between masses, for the sake of simplicity. Everything I say about the former equation can be framed in terms of the latter.
Which shows that the mass of the particle composed of particles 1 and 2 is the sum of the masses of particles 1 and 2. But the overlap simplification is strange. In Newtonian mechanics particles do not overlap in this way, they bounce off each other like billiard balls. Furthermore, equation (22) fails to describe the value of the field at the location of the particle determining that field (the self field). For this would involve a division of zero by zero. The gravitational relation between the overlapping particle is therefore not defined either. This is why I think explanations of gravitational mass additivity cannot be found in the physics literature: physicists use the inertial explanation because parts having identical accelerations is a genuinely physical phenomenon, whereas overlapping Newtonian particles is arguably unphysical. But then it is not clear that the general principle of Newtonian mass additivity has been explained by Kibble and Berkshire.

The problem is that when we apply our equations to a single entity, we require that the position and acceleration of that entity be described as the value of a single point in space. This is a problem because the positions and accelerations of composite objects are not generally defined at points. How can we solve this problem, without giving up on mass additivity?

In the next section I show how inertial mass additivity can be deduced from Newtonian fundamentals. In the section after that, I do the same for gravitational mass additivity. In the remainder of this section, I discuss a possible solution to the problem of extending the inertial argument to non-rigid bodies (i.e. composites with distinct component accelerations). This solution will come naturally to anyone used to thinking of composite position in terms of centre of mass.

The centre of mass solution involves the following strategy: Define the position of a composite at a point, and the acceleration of a composite (whose parts have distinct accelerations) as the acceleration of that point. Then, once one knows the acceleration and force of the composite, one can calculate inertial mass using the above procedure. This solution is defended in detail by Feather whose goal “is to derive a result valid without approximation” (1965, p511). However, Feather introduces the centre of mass abruptly, and does not defend its introduction. Following the textbook strategy, Feather considers three particles m₁, m₂, and m₃, located at distinct positions A, B, and C, and then provides three equations corresponding to equations (14), (15), and (16) from above. He then says:

“Now Let D be the point in BC such that \(m₂BD = m₃DC\). Then D is the instantaneous position of the centre of mass of the particles at B and C.”

---

44 The closest thing to such an explanation would be Newton’s Shell Theorem, which entails that a spherically symmetric body affects external objects gravitationally as though all of its mass were concentrated at a point at its centre. But this is not normally considered an explanation of gravitational mass additivity. However, as we will see in section 2.5, it is at least relevant to such an explanation.
The acceleration of point D is then crucial to Feather’s derivation. But no argument is given for introducing the point in BC (in the straight line connecting points B and C) that satisfies \( m_2 BD = m_3 DC \). This is not a derivation of mass additivity from microphysics alone. It is a derivation of mass additivity from microphysics plus an unmotivated stipulation about what the composite's position is.

Can there be a non-arbitrary, non-question begging argument for treating composite position as centre of mass? I have two arguments to the contrary: a metaphysical argument and a technical argument. The metaphysical argument is simple and (I believe) decisive and just involves the claim that it is metaphysically and conceptually impossible for a composite to be located at a single point, if the composite’s parts are located at a number of distinct points. For such a composite is patently located at a region (e.g. a set of distinct points), not a single point.

Many physicists will be attracted to the centre of mass solution. For when calculating the motion of a composite object, it is standard to calculate the motion of the composite’s centre of mass, which at any given time, is a single point in space. Arguably, this gives some kind of significance, metaphysical or otherwise, to this particular point in space. And so perhaps there is a non-arbitrary non-ad hoc way of treating the composite (and its mass) as being located at that point, and of treating the acceleration of the composite as the acceleration of that point.

In response, the technical argument makes two claims. Firstly, the significance of the centre of mass point is only due to the fact that Newton’s laws applied to that point can be used to calculate composite motion. Secondly, the proof that the laws apply to the centre of mass, uses composite mass as an axiom, and therefore presupposes mass additivity.

Firstly, consider why a composite's centre of mass might be thought to yield a solution. Recall equation (20), where we identified the acceleration of the composite with the acceleration of its parts such that: \( a_c = a_1 = a_2 \). We can still recover equation (20) if we assume \( a_2 \neq a_3 \), provided that the acceleration of the composite is the acceleration of the centre of mass:

\[
a_{cm} = \frac{d^2}{dt^2} \left( \frac{m_1 x_1 + m_2 x_2}{m_1 + m_2} \right)
\]

The second fraction represents the position \( x_{cm} \) of the centre of mass, defined by dividing the sum of the (mass weighted) component positions by the sum of the component masses. The first fraction signifies that we are taking the second time derivative of the second fraction, thus calculating the acceleration \( a_{cm} \) of the centre of mass. It turns out that equation (20) is equivalent to equation (27):

\[
F_c = (m_2 + m_3)a_{cm}
\]
So it looks like we can deduce mass additivity as before, in a way that applies to composites whose parts have distinct accelerations, provided that we have some non-arbitrary, non ad hoc way of showing that composite position is the centre of mass. Unfortunately, the only explanation I know showing that \( a_{cm} \) is even relevant to \( a_c \) (let alone metaphysically identical) explicitly presupposes mass additivity. A paradigmatic explanation is provided by Marc Lange:

“Let us see why it is the case in classical physics that Newton’s second law [...] “scales up” in that it governs the motions not only of the elementary bodies, but also of the centre of mass of a system of those bodies. [...] For Newton’s second law to govern this system’s motion would be for the force exerted on the system (namely, the sum of the forces exerted on the system’s constituents [...] to equal the system’s mass \((m_2 + m_3; \text{in classical physics, mass is additive})\) multiplied by the acceleration \(a\) of its centre of mass.”

(Lange 2002: 234 [my italics])

The italicised aspect shows that the explanation of the significance of the centre of mass presupposes mass additivity. In particular, if we know composite force and mass then we can show that centre of mass acceleration times composite mass equals the composite force. But if we don’t already know composite mass, we have no obvious reason to privilege the centre of mass. If we only know composite force, then different choices for “the centre” will yield different results for composite mass. For example, if we chose the following definition for the (point) position of the composite:

\[
x_c = \frac{\sum_i m_i x_i}{\sqrt{\sum_i m_i^2}}
\] (28)

Then by this method we would find that composite mass would have to be:

\[
m_c = \sqrt{\sum_i m_i^2}
\] (29)

Of course the definition in (28) is highly unnatural: the composite position depends on the choice of origin.45 So if one is able to respond to the metaphysical argument, then perhaps one could further motivate the claim that centre of mass is the most natural. But this looks unlikely as there are plenty of equally natural candidates such as the average position of the parts, or position of the most central part, or the most massive part, and so on. All of these will give slightly different results for the composite mass and so the centre of mass solution does not work.

---

45 To see this, let two one-dimensional particles have unit mass, and be located two meters apart at -1 and +1 respectively so that the origin is at 0. In that case (28) puts the composite at 0. But move the origin so the coordinates of the particles are now 99 and 101 and (28) puts the composite at the far away location 200√2.
It is therefore not surprising that in relativity, a different centre of mass definition is required. As Lange notes:

“In relativity theory, a collection of bodies likewise behaves as a single body with the system’s total mass m located at the collection’s centre of mass. A system’s centre of mass is again a weighted average of its constituents’ positions – but relativity weights them by their energies.”

(Lange 2002: 235).

Hence, our choice of centre of mass as composite position presupposes that we already know the composite mass. So composite position cannot be identified with centre of mass to reductively explain composite mass.

I now move to a different approach. This approach argues that alternative hypotheses about composite mass can be ruled out a priori, by appeal to microphysical simulations in which composites have parts nonuniformly accelerated by forces. Centres of mass play an important role, but at no stage are they identified with composite position. I use an a priori entailment expansion to express this approach and to show that inertial mass additivity follows a priori from Newtonian microphysics.

2.4 Inertial Mass

In this section I defend an a priori entailment expansion for the following conditional:

(Inertial mass additivity): If [Newtonian microphysical description] & [T] then inertial mass is additive.

I expand this conditional in accordance with the algorithm discussed in section 1.5. Let's begin by considering an instance of the antecedent that is simple but generalisable. We may confine ourselves to a microphysical description of a Newtonian world time slice containing only two elementary point masses, distinctly located but with identical instantaneous accelerations A.

We require terminology to describe both the composite of the two particles, and composites in nomologically possible situations: situations involving different initial conditions but the same Newtonian laws. Let C refer to the composite of the two point masses. Due to its identical component accelerations, C is a uniformly accelerating body or uniac body for short.46 Let C* refer to the

46 “Rigid body” would be inappropriate here as it refers to composites whose parts have interatomic forces that are large compared to external forces such that the particles are kept at a fixed distance from one another. If a
composite of the same two masses in any nomologically possible situation in which they have slightly distinct accelerations. C* is a *near-uniac body*. Let C** refer to the composite of the same two masses in any nomologically possible situation in which they have neither identical nor slightly distinct accelerations. C** is an *arbitrary fusion*. These three types of composite must be distinguished because the arguments for their additive masses are distinct. Finally, recall (from section 2.2) that 'playing an object's inertial mass role' means being the most natural property responsible for the object's disposition to resist changes in motion given applied forces. In this section I refer to it as the 'mass-role'. Every instance of 'mass' in this section refers to inertial mass.

I begin by setting out the expansion of (Inertial mass additivity). I then defend the apriority of each of its expansion conditionals. Note that each expansion conditional (premises (2)-(8)) is intended to represent an a priori inference from fundamental Newtonian microphysics to some non-fundamental proposition. It is important that the antecedents of the expansion conditionals appeal not only to premise (1), but also to premises that precede them. This is because the consequents of such conditionals may not be derivable from (1) without first deriving certain other information from (1).

Here is the expansion:

1) **[Newtonian microphysical description]** There are three point masses m1, m2 and m3, located in Euclidean space at distinct positions x1, x2, and x3, such that m1 and m2 have identical accelerations A. The force F on m3 due to m1 and m2 is determined by Newton’s second law and the superposition of forces:

$$F_i = \sum_{j=1}^{N} F_{ij}$$

That’s all that is fundamental.

2) **[Composition]** If (1) then there is a composite C composed of m1 and m2 located where m1 and m2 are located (the set of points \{x1, x2\}), in virtue of m1 being at x1 and m2 being at x2.

3) **[Force]** If (1) & (2) then there is a force F on C, in virtue of the force F on C's parts.

4) **[Acceleration]** If (1) & (2) then C has acceleration A, in virtue of its parts each having acceleration A.

A composite is rigid in this sense and it is not rotating about its centre of mass then the accelerations of the parts of a rigid body will be equal. But the converse does not hold. As far as I know there is no expression for a composite whose parts have identical accelerations so I introduce the notion of a uniac body.
(5) **[Uniac Mass]** If (1)-(4) then C’s mass-role is being played by \([m_1+m_2]\) in virtue of \(F=A[m_1+m_2]\). So mass is additive for uniac composites.

(6) **[Near-Uniac Mass]** If (1)-(5) then C*’s mass-role is played by \([m_1+m_2]\) in virtue of \(F=A*[m_1+m_2]\), where A* is a member of the admissible precisifications for C*’s acceleration. So mass is additive for near-uniac composites.

(7) **[Arbitrary-Fusion Mass]** If (1)-(6) then C**’s mass-role is played by \([m_1+m_2]\) in virtue of \([m_1+m_2]\) playing C’s mass-role and C*’s mass-role. So mass is additive for arbitrary fusions.

(8) **[Mass Additivity]** If (1)-(7) then mass is additive.

(9) **Macrophysical description**: mass is additive.

If (2)-(8) are a priori then the conditional 'If (1) then (9)', and hence our target conditional (Inertial mass additivity), is a priori. I shall now defend the apriority of each expansion conditional in turn.

**[Composition]**

The most straightforward defence of the apriority of [Composition] appeals to the Analysis Based Justification heuristic from section 1.2. Our concept of a composite object applies just in case there are some objects that can in principle be grouped in thought and named. Thus, one only needs this ability (grouping and naming) to be in a position to infer all composites, from knowledge of what elementary particles exist. One can therefore defend the apriority of [Composition] by appeal to the application conditions that are defining of the term 'composite' that is used to state [Composition]. While some philosophers might deny that this concept is the one that ought to be expressed by ‘composite’, they surely cannot deny that such a concept is coherent and is an admissible disambiguation of one concept sometimes expressed by ‘composite’. For example, when physicists illustrate mass non-additivity in relativity theory, by quantifying over the object composed of two photons travelling in different directions, it is hard to see what other notion of composite they could be using (Okun 2008: 626). Furthermore, while most philosophers accept unrestricted composition, there is clearly no experience that confirms unrestricted composition over restricted composition. So arguably, this widespread belief is based not on what the actual world is like, but on the application conditions for the concept itself. So for present purposes, we can argue that there is a legitimate notion of a composite that initiates our a priori entailment expansion for (Inertial mass additivity). It also needs to be argued that this notion of a composite is an important one and that the existence of
alternative notions do not undermine the significance of the present argument. To maintain the flow of the defence of the a priori entailment expansion, I develop these arguments in detail in section 2.8.

[Force]

Recall from the textbook explanation that we may derive: \( m_3a_3 = -m_1a_1 - m_2a_2 \). Using the second law we may then derive: \( F_3 = -(F_1 + F_2) \). So the force on particle 3 is equivalent to the sum of the forces on particles 1 and 2, oppositely directed. From this we must derive: \( F_3 = -(F_0) \) (the *additivity of forces*). I shall do this in four steps. Step 1 derives the claim that C exerts some force on particle 3. Step 2 derives the claim that that force must be \(- (F_1 + F_2)\). Step 3 derives the claim that particle 3 must exert some force on C. Step four derives the claim that that force must be \( F_1 + F_2 \).

Step one: to derive the claim that C exerts a force on particle 3 our deflationary analysis of composition will be essential. Necessarily, composition holds if and only if grouping and naming is an in principle possibility. Hence, necessarily, removing C from the situation described in premise (1) removes at least one of the two particles whose grouping and naming enables one to deduce C. This removes force \( F_2 \) and/or force \( F_1 \). But this removes force \( F_3 \). So C is not “forceless” because necessarily, if we remove it from the situation in (1) the force on particle 3 changes.

Step two: we now derive the value of the force that C exerts on particle 3. There can only be one value and we can prove by reductio that this value is \( F_3 = -(F_1 + F_2) \). Assume that C exerts a force on particle 3 that is not equal to \( F_3 \). Since we know that particles 1 and 2 exert force \( F_3 \) on particle 3 and since we know that \( F_3 \) is the only force on particle 3, we can infer that the force that C exerts on particle 3 must be identical to the force that C’s components exert on particle 3. And so we may derive \( F_{3C} = -(F_1 + F_2) \), where \( F_{3C} \) is the force on particle 3 due to C.

Step three: to derive the claim that particle 3 exerts a force on C we appeal to considerations similar to those used in step one. If we remove particle 3 we thereby affect C’s acceleration. Thus, from the fact that removing (or even changing the state of) particle 3 affects C in this way we are justified in postulating a force on C due to 3. After all, forces are postulated to explain accelerations. The question is what the value of this force is.

Step four: by stipulation the relevant notion of force is governed by a naturalness constraint, as are all the notions relevant to the proof. As discussed in section 2.2, a term is governed by a naturalness constraint if and only if where possible it applies to properties (that play some conceptual role) that are *most like* the fundamental properties. The idea is that similarity to the fundamental properties acts as a kind of reference magnet for higher-level terms. We know from premise (1) that fundamental forces have the property of obeying Newton’s third law. In that case, given that \( F_{3C} = -(F_1 + F_2) \),
the sum of the forces of C’s parts becomes the best candidate extension for ‘the force on C’. One is therefore a priori justified in deducing $F_3 = -F_c$.

We now go back to $F_3 = -m_1a_1 - m_2a_2$, and substitute:

$$F_c = m_1a_1 + m_2a_2$$  \hspace{1cm} (30)

There are considerations in the metaphysics literature that might be used as a critique. Mereological nihilists (who think objects never compose) may reject [Force] by reasoning as follows: if there was a force on C then it would over-determine the force on the parts. Over-determination is bad, so there can be such force. Furthermore, assuming that such ‘forceless epiphenomena’ are bad the nihilist may deny the existence of the composite, thus denying [Composition] too. In response, when I say that we can infer from (1) and (2) that there is a particular force on C, I am not saying that this force is something over and above the force on C’s parts. Rather, the force on C just is the force on its parts. This is partly why physicists use the composite’s force to make inferences about the total force on the parts. This avoids the over-determination concern.

[Acceleration]

If we restrict ourselves to a situation in which the components have identical accelerations, then we can deduce composite acceleration. For in this situation (a uniac situation), composite acceleration is just the acceleration shared by its components. Thus [Acceleration] only requires the apriority of the following principle:

**Restricted Acceleration Composition** ($A_c = A_j$): composite acceleration is the acceleration of one of its parts, when the accelerations of its parts are identical.

The negation of this principle is inconceivable: there is no other value for composite acceleration that makes sense. But the composite is accelerating if its parts are accelerating. So composite acceleration is a priori deducible.

[Uniac Mass]

Equation (30) can be written more generally (particle 3 interacting with only two other particles was a mere artefact of the example):

\[ This line of thought mirrors Trenton Merricks’ (2001: 80-81) argument for mereological nihilism in which the baseball’s supposed shattering of the window overdetermines the atoms’ shattering of the window. See also Wilson (2010).

\[ Amie Thomasson (2006) provides an illuminating response along similar lines.\]
Because component accelerations are identical, a simple algebraic transformation takes the acceleration term out of the scope of the sum:

\[ F_c = \sum_{i=1}^{N} a_i m_i \]  

(31)

The subscript ‘i’ has been replaced with 1, designating component particle 1. Since all component accelerations are identical to \(a_1\) we appeal to Restricted Acceleration Composition to derive the following result, which relates composite force and acceleration, to component masses:

\[ F_c = a_1 \sum_{i=1}^{N} m_i \]  

(32)

From here, we only need to appeal to our a priori definition of mass:

**Definition of Inertial Mass**: The mass of an object is a measure of the object’s most natural disposition to resist changes in motion given applied forces.

Given equation (33), we know that the sum of the component masses is, as a matter of natural law, the coefficient relating composite force to composite acceleration. In particular (33) entails that C’s changes in motion given a force is proportional to the sum of the component masses. Hence, we can infer that the composite’s (most natural) disposition to resist changes in motion given applied forces just is the sum of the component masses. Hence, mass is additive for uniac composites.

**[Near-Uniac Mass]**

We have only derived a restricted result. What about composites with distinct component accelerations? We have not ruled out a priori the hypothesis that composite mass changes (becomes more/less massive) when component accelerations become distinct. On this hypothesis, mass is not additive.

Extreme versions of this hypothesis can be ruled out a priori. Imagine C is stationary, until total force \(F\) is applied, yielding acceleration \(A\) (that is, yielding acceleration \(A\) for both components). Let \(A = 1000 \text{m/s}^2\) in the direction of the applied force. Now take \(C^*\), which is identical to C in every respect except \(C^*\) is not quite stationary: component 1 is stationary while component 2 begins with an imperceptibly small acceleration (in the direction of the force we are about to apply to \(C^*\)). We then apply force \(F\) to \(C^*\), just as we did to C. This might involve a collision experiment which fires a two
particle system $S$ at $C$, and the same two particle system $S$ at $C^*$. We apply the same force $F$ in the sense that the total force on $S$ is the same each time. The resulting acceleration of $C$ in response to $F$ is $1000\text{m/s}^2$ in the direction of $F$. But imagine the resulting acceleration of $C^*$ in response to $F$ is quite different, in that $C^*$ barely moves, acquiring only an extremely small resulting acceleration. Thus, while the force has a significant effect on uniac body $C$, it barely moves near-uniac body $C^*$. Such a situation would be evidence against mass additivity. For here, composite disposition to resist acceleration given a force (i.e. a composite mass) is a function, not only of component masses, but of component accelerations too.

It is unclear exactly what it means for a near-uniac body to “barely move, acquiring only an extremely small resulting acceleration”. That’s because we cannot attribute an exact value to the resulting acceleration. However, we can attribute exact resulting accelerations to components. And in the extreme hypothesis, $C^*$ components react very differently to $C$ components in response to $F$, despite having the same masses. Because of this, Newton’s laws cannot be the complete fundamental microphysical laws (as stated in premise (1)). A more complex law than $F=MA$ would have to obtain in order for the extreme hypothesis to obtain. For example, $F=MAO$, where $O$ is some other factor, which constrains the components of $C^*$ in response to $F$, but allows the components of $C$ to accelerate in response to $F$. Since the relevant fundamental law is $F=MA$ rather than something like $F=MAO$, we can rule the extreme hypothesis out, a priori.

There are still more conservative hypotheses that we are yet to rule out. The above considerations do not force us to conclude that near-uniac composite mass is just whatever the corresponding uniac composite mass is. (Here ‘corresponding’ entails ‘same component masses’.) For perhaps near-uniac composite mass is slightly distinct from the corresponding uniac composite mass. This conservative hypothesis is not inconsistent with the claim that Newton’s laws (e.g. $F=MA$ rather than $F=MAO$) are the complete microphysical laws. There is room for this conservative hypothesis because of the vagueness of the notion of ‘composite acceleration’ when applied to a near-uniac body.

One response treats composite mass as vague. The idea is that if by ‘mass’ we mean a measure of the disposition to resist acceleration, but acceleration resistance is vague, then mass is equally vague. This is a legitimate mass concept. But it does not appear to be our mass concept. For physicists readily attribute precise masses to precise composites (i.e. composites with determinate components) even when the components have distinct accelerations. This attribution can be explained by the naturalness constraint on ‘mass’ discussed in section 2.2. Applying ‘mass’ in accordance with the naturalness constraint means applying it (where possible) to a property that is like a fundamental property; and fundamental properties are not vague, at least not in Newtonian worlds. Let us then consider why additive mass is the most natural mass for near-uniac bodies in this sense.
Our fundamental properties are all precisely specified. So a plausible naturalness constraint on 'mass' is what I call the “best precise deserver constraint”: where possible apply ‘mass’ to whatever precise value is the best measure of an object’s disposition to resist acceleration given applied forces. Furthermore, while our fundamental kinematical properties (position, acceleration) are constantly changing, our fundamental dynamical properties (mass) are constant. So a second plausible naturalness constraint on 'mass' is what I call the “uniformity constraint”: where possible apply ‘mass’ to whatever constant value is the best measure of an object’s disposition to resist acceleration given applied forces. Each constraint has a where possible clause. This just means that if there is no such precise/constant value, ‘mass’ still applies if something plays the mass-role.

Because we are considering near-uniac bodies (rather than arbitrary fusions) the vagueness of the composite acceleration is minimal. Thus, the set of admissible precisifications for C*’s acceleration is quite constrained. Here, an admissible precisification of C*’s acceleration is a precise acceleration that satisfies all the facts about C*’s vague acceleration, that can in principle be satisfied by a precise value. To determine admissible precisifications one can simulate a collision experiment that accelerates C*, and accelerate uniac bodies (or better, point particles) alongside C*. We can then infer many facts about C*’s acceleration: C* is accelerating faster than test particle P but slower than test particle P*. To deduce the set of admissible precisifications for C*’s acceleration, we take the set of precise accelerations that satisfy these claims. We end up with a highly constrained continuum of point values. This is useful, because we already know what C*’s force is. We can then use the set of precise accelerations, together with the second law, to determine which precise mass value best represents the composite’s disposition to resist acceleration given the force. We can draw up the following table:

<table>
<thead>
<tr>
<th>Admissible acceleration precisifications</th>
<th>Mass hypothesis</th>
<th>Second law calculations</th>
</tr>
</thead>
</table>
| \{A_1, A_2, A_3, A_4, ... A_N\}        | M_U (additive uniac mass) | F_{C*} = M_U A_{COM}  
F_{C*} \sim M_U A_1  
F_{C*} \sim M_U A_2  
...  
F_{C*} \sim M_U A_N |
| Same as above                           | M_{U+1} (near additive mass) | F_{C*} \sim M_{U+1} A_1  
F_{C*} \sim M_{U+1} A_2  
...  
F_{C*} \sim M_{U+1} A_N |
The point of the calculations in the third column is to determine which precise mass candidate is the most accurate measure of C*'s disposition to resist acceleration given applied forces. The first hypothesis is that C*'s mass is identical to C’s mass (the additive mass we deduced for uniac bodies). It turns out, that there is an admissible precisification of C*'s acceleration that is equivalent to C*'s force divided by M_U: the acceleration of C*'s centre of mass, A_COM.49 Furthermore, because every other admissible acceleration precisification is very close to A_COM, it follows that the product of M_U and each precisification is very close to the actual force on C*. So M_U becomes a clear candidate for being the most natural mass-role player for C*.

Our two naturalness constraints on ‘mass’ are individually sufficient to guarantee that ‘mass’ applies to M_U and to no other value. Regarding the best precise deserver constraint: as we increase the mass hypothesis (e.g. M_U+1) or decrease the mass hypothesis (e.g. M_U-1), we find that the second law calculations slowly become less and less approximately true. To see this (without doing all the calculations) note that as the mass hypothesis gets higher (or lower) equality eventually falls out of the third column of the table. That is, no longer is there an admissible precisification that is equal to the actual force divided by the mass hypothesis. Thus, we can rule out mass hypotheses that do not equal force divided by any one admissible acceleration precisification. For they are inaccurate measures of acceleration resistance given force. And we can conclude that M_U is the most accurate precise measure because dividing it by the admissible acceleration precisifications yields a set of values that approximate the force better than any other precise mass hypothesis.

The uniformity constraint pressures us to apply ‘mass’ uniformly to uniac and near-uniac composites, unless there is empirical reason not to. So even if a mass hypothesis close to M_U better approximated to the force multiplied by the acceleration precisifications, this would not be sufficient to make it a better candidate mass than M_U. For insofar as the candidate is only slightly better than M_U at approximating the force, the uniformity constraint would win out.

I have argued that there is a concept of mass, which works in accordance with the constraints discussed above, which is at least not far from our mass concept, and which guarantees that mass additivity is a priori derivable for near-uniac bodies. It is plausible that we accept mass additivity for near-uniac bodies because our mass concept is the mass concept I have described. So mass is additive for near-uniac bodies.

[Arbitrary-Fusion Mass]

We generally attribute mass to composites no matter their configuration. So one might agree with everything so far, but deny that we have a priori inferred mass additivity by arguing that we have not

---

49 Recall from section 2.3 that for a composite with two components, with masses m_1 and m_2 with respective positions x_1 and x_2, the centre of mass is the point in space determined by: (m_1x_1 + m_2x_2) / (m_1 + m_2).
ruled out the possibility that when composites become highly disconfigured (so that collision experiments are no longer meaningful), mass additivity fails. Call such a composite C**. Now, insofar as C**’s mass cannot manifest at all; that is, insofar as there is no in principle test of C**’s mass, then its mass is just whatever we find it to be in uniax and near-uniac scenarios. This result is guaranteed by the uniformity constraint. There is no empirical reason to treat C**’s mass as an unnatural property that varies between C/C* and C** scenarios, so the uniformity constraint determines C**’s mass given the mass of C and C*. Hence, m1+m2 plays the mass role for C**.

We can now derive the general result that C’s mass is additive. The argument can be run for all but one of the composites implied by the microphysical description in premise (1). The composite it leaves out is the one composed of all three particles. But we are justified in deducing mass additivity for the global composite by introducing a hypothetical test particle into the world and then running the argument. We may then infer that mass is additive in the world described in (1). And note that a world containing only three particles was chosen for simplicity. We could have used any Newtonian composite with arbitrarily many components and the argument would still go through just the same. Inertial mass additivity therefore follows a priori from Newtonian microphysics. Let us now consider gravitational mass additivity.

2.5 Gravitational Mass

In this section I defend an a priori entailment expansion for the following conditional:

(Gravitational mass additivity): If [Newtonian microphysical description] & [T] then gravitational mass is additive.

We may confine our antecedent to a microphysical description of a time slice of a Newtonian world containing only two elementary point masses. This time, they have identical positions and therefore perfectly overlap. This may seem like an odd requirement. But if we are dealing with a world that does not obey any collision laws (let alone Newton's) but only obeys Newton's law of gravity, then given enough time, overlap is what we should expect. A more severe problem is the fact that Newton’s gravitation equation is not suited to describe the value of the field at the position that the particle determining that field is located at (the self-field). So it cannot describe the gravitational interaction between two overlapping point particles. This is because the equation requires that one

50 Note that an alternative is to deny that arbitrary fusions exist. In section 2.8 I show how AET works under the assumption of restricted composition.
divide by $x - x_i$ (see equation (34) below) where $x_i$ is the position of the particle and $x$ is the position corresponding to the field value we are interested in. If we are interested in the field value for the position at which the particle is located, then the equation requires that we divide zero by zero. There are ways of fixing this (e.g. changing the equation to guarantee that the value of the field at the particle’s location is zero). Instead, I just note that the physicality of the overlap situation is not required for my argument, but is only required to define a certain kind of “ideal case” composite whose behaviour can be compared to that of “compact” composites.

We require terminology similar to the previous section, to describe both the composite of the two overlapping particles in the described situation, and composites of the same parts in nomologically possible situations where they don’t overlap. Let C refer to the composite of the two point masses. Due to its identical component positions, C is an overlap body. Let C* refer to the composite of the same two point masses in any nomologically possible situation in which they have slightly distinct positions. C* is a compact body. Let C** refer to the composite of the same two point masses in any nomologically possible situation in which they have neither identical nor slightly distinct positions. C** is an arbitrary fusion. Finally, recall (from section 2.2) that 'playing an object's gravitational mass role' means being the most natural property responsible for the object's disposition to attract other objects as they come closer. In this section I just refer to the 'mass-role'. Every instance of ‘mass’ in this section refers to gravitational mass.

Here is the expansion:

(1) **[Newtonian microphysical description]** There are two point masses $m_1$ and $m_2$ both located at the same position $x_1$ in Euclidean space. The gravitational force field $g(x)$ determined by the two masses is given by Newton’s gravitation law and the composition of fields:

$$g(x) = G \sum_{i=1}^{N} \frac{(x - x_i)}{|x - x_i|} m_i$$

That’s all that is fundamental.

(2) **[Composition]** If (1) then there is a composite C composed of $m_1$ and $m_2$ located where $m_1$ and $m_2$ are located (point $x_1$), in virtue of $m_1$ being at $x_1$ and $m_2$ being at $x_1$.

(3) **[Field]** If (1) & (2) then C determines gravitational field $g(x)$, in virtue of C's parts determining $g(x)$. 


(4) **[Overlap Mass]** If (1)-(4) then C’s mass-role is being played by \([m_1+m_2]\) in virtue of \(g(x) = G \left[ \frac{(x-x_1)}{|x-x_1|} (m_1 + m_2) \right]\). So mass is additive for overlapping composites.

(5) **[Compact Mass]** If (1)-(5) then C*’s mass-role is played by \([m_1+m_2]\) in virtue of \(g(x) = G \left[ \frac{(x-x^*)}{|x-x^*|} (m_1 + m_2) \right]\), where \(x^*\) is a member of the admissible idealisations for C*’s position. So mass is additive for compact composites.

(6) **[Arbitrary Fusion Mass]** If (1)-(6) then C**’s mass-role is played by \([m_1+m_2]\) in virtue of \([m_1+m_2]\) playing C’s mass-role and C*’s mass-role. So mass is additive for arbitrary fusions.

(7) **[Mass Additivity]** If (1)-(8) then mass is additive.

(8) **Macrophysical description**: mass is additive.

If (2)-(7) are a priori then the conditional 'If (1) then (8)', and hence our target conditional (Gravitational mass additivity), is a priori. I shall now defend the apriority of each expansion conditional in turn.

**[Composition]**

As this premise is effectively identical to the composition premise from the previous section, I refer the reader to that discussion, as well as to the more detailed defence, in section 2.8.

**[Field]**

Here we put the composition of fields principle (from section 2.3) together with [Composition], to infer what I call the additivity of fields:

**The additivity of fields** \((g(x) = g(x)_c)\): If there is a total gravitational field \(g(x)\) determined by a group of particles that compose composite C, then C determines \(g(x)_c\).

We know a priori that if the composition of fields holds for particles that compose a composite C then the field determined by the composite is equivalent to the field determined by the parts. Imagine experimentally testing the additivity of fields. We cannot falsify the principle for the following reason: were we to find a distinct value for the composite field, we would automatically attribute that field to the components, thus correcting our previous estimate of the field determined by the components. Experimentally testing the additivity of fields makes no sense, which is evidence that there is no conceivable world that falsifies the additivity of fields, which is evidence that it is a priori.
If this principle is a priori then we can infer the field \( g(x)_C \) determined by C and how \( g(x)_C \) is determined by the masses and positions of C’s parts:

\[
g(x)_C = G \sum_{i=1}^{N} \frac{(x - x_i)}{|x - x_i|^3} m_i
\]  
(35)

[Overlap Mass]

Because the positions of C’s parts are identical, a simple algebraic transformation enables us to take the fraction out of the scope of the sum in (35) and get:

\[
g(x)_C = G \frac{(x - x_c)}{|x - x_c|^3} \sum_{i=1}^{N} m_i
\]  
(36)

The subscript ‘i’ has been replaced with C: as [Composition] states, the position of the composite just is the positions of the parts, and here, that amounts to a single point in space.

From here, we only need to appeal to our a priori definition of mass:

**Definition of Gravitational Mass**: The mass of an object is a measure of the object’s most natural disposition to attract other objects as they come closer.

Equation (36) entails that the sum of the component masses is, as a matter of natural law, the coefficient relating composite field to composite position relative to that field. In particular, (36) entails that the extent to which C attracts objects in its field is proportional to the sum of the masses of its parts. Hence, we can infer that the composite’s (most natural) disposition to attract other objects as they come closer just is the sum of the masses of its parts. Hence, mass is additive for overlap composites.

[Compact Mass]

We only derived a restricted result. What about the masses of composites whose parts have non-identical positions? We have not ruled out a priori the hypothesis that composite mass changes (becomes more/less massive) when component positions become distinct. On this hypothesis, mass is not additive.

As in the inertial case, extreme versions of this hypothesis can be easily ruled out a priori. For example, imagine our overlap composite determines some field \( g(x) \), until its parts become slightly separated in space, which induces the field to become immensely strong, immediately pulling everything towards the region it inhabits. This situation is inconsistent with the microphysical laws, because of what is required to allow the components to determine such a field. So we can rule out the
extreme hypothesis by noting that there is nothing in the underlying laws that makes the field a function of the relative positions in this extreme way. In particular, by simulating the situation in which C (overlap composite) transforms into C* (compact composite), we can deduce that the resulting field does not change much at all. The extreme hypothesis is therefore ruled out a priori.

What are not ruled out a priori by these considerations are more conservative hypotheses stating that mass changes slightly during the transformation from an overlap body to a compact body (or vice versa). Here, there is an important difference with the inertial case. In the inertial case, the conservative hypotheses thrived on the vagueness of the notion of composite acceleration resistance. For this reason, it was not enough to just appeal to the microphysics to rule the conservative hypothesis out. One also needed to appeal to semantic constraints and admissible precisifications of composite acceleration. But in the gravitation case there is no vagueness. Composite position is entirely determinate—a compact composite is located in a region. More precisely, the composite is located where its parts are located (a set of positions).

The conservative hypotheses therefore do not thrive on vagueness. Rather, they thrive on the fact that the gravitational field is affected slightly by C’s transformation from an overlap composite to a compact composite, because the total field is a function of component relative positions. So what rules out the hypothesis that composite mass changes slightly during such a slight transformation in configuration?

Firstly, our naturalness constraints on the application of ‘mass’, will play a key role. I argue that each constraint—the best precise deserver and uniformity constraints—are each individually sufficient to determine the right additive result. But these constraints cannot interact with “admissible precisifications” because there is no relevant vague predicate to make precise. The constraints can, however, interact with admissible idealisations in order to derive the desired result. The idea is that the position of a compact body can be idealised as a constrained set of exact points in space such that if we put those individual points into the gravitation equation, along with additive mass, we recover the composite field, at least to a higher degree of approximation than any other mass hypothesis.

Admissibility is defined as it was in the inertial case. In the inertial case, an admissible precisification of C*’s acceleration was a precise acceleration that satisfies all the facts about C*’s vague acceleration, that can in principle be satisfied by a precise value. Here, an admissible idealisation of C*’s position is a point in space that satisfies all the facts about C*’s position, that can in principle be satisfied by a point location. The most relevant facts are of the form ‘is found within region R’. We end up with a highly constrained continuum of point locations. This is useful, because we already know what C*’s field is. We can then use the set of point locations, together with the gravitation law,

51 This is not to say that the position of a compact composite is an abstract entity. One should not reify my talk of sets, as set theoretic vocabulary is just a convenience here. I could just as well define the composite’s position as these points (pointing at specific points in the simulation of Euclidean space).
to ask which mass value best represents the composite’s most natural property responsible for its disposition to attract other objects as they come closer.

Why let facts about admissible *idealisations* determine the exact mass of the composite being idealised? The conservative hypothesis is thriving on the fact that the field determined by a composite is determined, not just by the masses of its parts, but also by its internal structure: the relative positions of its parts. Idealising a compact body as a single point mass allows us to ask what mass determines the same field as the compact body, at least at distances at which the effects of the internal structure of the compact body are negligible.

Once we calculate the admissible idealisations, we can draw up the following table:

<table>
<thead>
<tr>
<th>Admissible position idealisations</th>
<th>Mass hypothesis</th>
<th>Gravitation law calculations</th>
</tr>
</thead>
<tbody>
<tr>
<td>{x_1, x_2, x_3, x_4, \ldots, x_n}</td>
<td>M_0 (additive overlap mass)</td>
<td>[ g(x)<em>{c^*} \sim G \left( \frac{x-x</em>{cog}}{</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[ g(x)_{c^*} \sim G \left( \frac{x-x_1}{</td>
</tr>
<tr>
<td></td>
<td></td>
<td>\ldots ]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[ g(x)_{c^*} \sim G \left( \frac{x-x_n}{</td>
</tr>
<tr>
<td>Same as above</td>
<td>M_{0+1} (near additive mass)</td>
<td>[ g(x)_{c^*} \sim G \left( \frac{x-x_1}{</td>
</tr>
<tr>
<td></td>
<td></td>
<td>\ldots ]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[ g(x)_{c^*} \sim G \left( \frac{x-x_n}{</td>
</tr>
</tbody>
</table>

The point of the calculations in the third column is to determine which mass candidate is the most accurate measure of C*’s disposition to attract other objects as they come closer. The first hypothesis is that C*’s mass is identical to C’s mass (the additive mass we deduced for overlap composites). It turns out, that there is an admissible idealisation of C*’s position—\(x_{cog}\)—such that when we put into the gravitation equation, together with the additive overlap mass, we recover the composite’s field to the highest possible degree of accuracy. This is the position of the composite’s *centre of gravity*. In particular, as you move further out from the location of the compact body, the field determined by the additive mass at that centre tends towards the field determined by the compact body. A diagram may help:
Let the two dots on the left hand image be two point masses and let the dot on the right hand image be a single point mass whose mass is the sum of the other two. The rings around the masses represent field lines, where the outer rings represent weaker parts of the field. The image illustrates how the gravitational field on the right becomes more and more similar to the total gravitational field on the left, the further out we go. This is despite significant differences in the interior, due to the internal configuration of the compact composite on the left.

Because every other admissible position idealisation is very close to $x_{\text{cog}}$ it follows that the other equations involving $M_0$ are very close to equality too. So $M_0$ becomes a clear candidate for the mass of C*. Our two naturalness constraints on ‘mass’ are individually sufficient to guarantee that ‘mass’ applies to $M_0$ and to no other value. Regarding the best precise deserver constraint: as we increase the mass hypothesis (e.g. $M_{0+1}$) or decrease the mass hypothesis (e.g. $M_{0-1}$), we find that the calculations slowly become less and less approximately true. On the other hand the uniformity constraint pressures us to apply ‘mass’ uniformly to overlap and compact composites, unless there is empirical reason not to. So even if a mass hypothesis close to $M_0$ better approximated the field when entered into the gravitation equations, this would not be sufficient to make it a better candidate mass than $M_0$. For insofar as the candidate is only slightly distinct from $M_0$, the uniformity constraint would win out.52

I have argued that there is a concept of mass, which works in accordance with the constraints discussed above, which is at least not far from our mass concept, and which guarantees that mass additivity is a priori derivable for compact bodies. It is plausible that we accept mass additivity for compact bodies because our mass concept is the mass concept I have described. So mass is additive for compact bodies.

[Arbitrary-Fusion Mass]

We generally attribute mass to an object no matter its configuration. So one might agree with everything so far, but deny that we have a priori inferred mass additivity by arguing that we have not ruled out the possibility that when composites become highly disconfigured (so that composite gravitational attractions are no longer meaningful), mass additivity fails. Call such a composite C**. It is not clear that there are such things. For it seems that no matter how disconfigured a composite is, it can exhibit meaningful gravitational attractions. Indeed, the further out you go, the more alike such attractions will be to the arbitrary fusion’s centre of gravity. So the above argument will apply to such composites. Insofar as there are arbitrary fusions so disconfigured that gravitational attractions can no

---

52 Newton’s Shell Theorem entails that a spherically symmetric body affects external objects gravitationally as though all of its mass were concentrated at a point at its centre (the cog). That is, the gravitational effect of a spherically symmetric massive body would, outside its limits, be identical to that of a point mass located at its centre whose mass is the sum of its parts. But note that this requires spherical symmetry so is limited in application.
longer be meaningfully attributed to the composite, the uniformity constraint will decide the matter for us, and C**’s mass will be additive.

We can now derive the general result that C’s mass is additive. And note that C was just a typical example. We could have used any Newtonian composite with arbitrarily many components and the argument would still go through just the same. Gravitational mass additivity, like inertial mass additivity, therefore follows a priori from Newtonian microphysics. This is the general result we were after. In the next section I consider an alternative argument for gravitational mass additivity. Then in section 2.7 I show that the arguments for gravitational mass additivity apply equally well to charge additivity. Then in section 2.8 I consider [Composition] in more detail. I then conclude the chapter with a more general defence of AET.

2.6 Gravitational Mass Density

This section is the most technical of the dissertation and can be skipped by readers satisfied by the previous section. It illustrates an alternative method for deducing gravitational mass additivity, which derives a more complicated formulation of the fundamental law before deriving mass additivity. Applying the inertial strategy (section 2.4) to gravitational mass (section 2.5) forces us to appeal to overlapping particles and admissible idealisations. If one is sceptical of these notions, one will be sceptical of the argument in section 2.5. It is therefore worth asking whether we can do without such notions.

Since the positions of the relevant composites are determinate sets of positions, perhaps we just need to formulate the gravitation equation as a function of sets of positions. An integral formulation will enable this. The integral formulation requires that we attribute mass densities to objects. We can then use the textbook method (section 2.3) to derive mass density additivity. The question will be whether mass density additivity is sufficient for mass additivity.

The density function used to describe a point mass located at \( x_0 \) is the Dirac delta function \( \delta \) defined as follows:

\[
\delta(x - x_0) = \begin{cases} 
\infty, & \text{if } x = x_0 \\
0, & \text{otherwise}
\end{cases}
\]

The Dirac delta function \( \delta \) therefore assigns \( \infty \) to the point \( x_0 \), and 0 to every other point in space. The integral of the Dirac delta function is defined so that if we integrate over it we get unity:
\[ \int \delta(x - x_0) \, dx = 1 \]

This is useful because we can then define the mass density \( \rho(x) \) of a point mass located at \( x_0 \) with mass \( m \) as follows:

\[ \rho(x) = m \, \delta(x - x_0) \]

And when we integrate over this function we get \( m \):

\[ \int \rho(x) \, dx = \int m \, \delta(x - x_0) \, dx = m \]

With these definitions we can formulate an equation that is equivalent to the original gravitation equation. Thus, for a single mass density:

\[ g(x) = G \int \frac{m \, \delta(\xi - x_0)(x - \xi)}{|x - \xi|^3} \, d\xi \]

This is equivalent to the original equation because the integral over a distribution of points is equivalent to a sum over the points in the distribution. So the integral is just playing the same role as the composition of fields. The generalised law for multiple mass distributions each with distinct masses \( m_i \) concentrated at distinct points \( x_i \), is then:

\[ g(x) = G \sum_{i=1}^{N} \left[ \int \frac{m_i \, \delta(\xi - x_i)(x - \xi)}{|x - \xi|^3} \, d\xi \right] \]

The fact that there are two sums in this equation (sigma and the integral) is simply an extension of the composition of fields. So this equation is derivable from Newton's equation from the previous section.

We may now apply the textbook transformation and take the position terms outside the scope of the sum by swapping the sum and the integral:

\[ g(x) = G \int \left[ \frac{(x - \xi)}{|x - \xi|^3} \sum_{i=1}^{N} m_i \, \delta(\xi - x_i) \right] \, d\xi \]

Since this recovers the form of the single particle (single density) law, it demonstrates the additivity of mass densities. For \( g(x) \) is the field determined by the composite of the \( N \) densities. So if \( \rho_c(x) \) is the composite of the densities indexed by \( i \), then:

\[ \rho_c(x) = \sum_{i=1}^{N} m_i \, \delta(x - x_i) \]
This is just the textbook argument performed with equations designed to apply to distributed composites.

But arguably, this is insufficient: our goal was not to derive the additivity of mass densities, but the additivity of mass. To derive the additivity of mass, we just need to motivate the idea that the mass of an object with a complex Dirac delta function is its integral. For the so-called sifting, or selector property of Dirac delta functions entails that the integral of a sum mass densities is the sum of the masses. The selector property is stated as follows:

\[
\int \sum_{i=1}^{N} f(x) \delta(x - x_i) dx = \sum_{i=1}^{N} f(x_i)
\]

The selector property can be proved as follows. Consider the following integral:

\[
\int f(x) \delta(x - x_0) dx
\]

Here \( f(x) \) is just some function of \( x \), for example, a function that assigns mass values to points in space. Now, since \( \delta \) is zero everywhere except at \( x_0 \), the entire integrand is zero everywhere except at that point. This means that the only contribution of \( f(x) \) to the integral happens at \( x_0 \), and it can therefore be written as:

\[
\int f(x) \delta(x - x_0) dx = \int f(x_0) \delta(x - x_0) dx = f(x_0) \int \delta(x - x_0) dx = f(x_0)
\]

If we’re considering the addition of multiple delta functions, we simply write:

\[
\int \sum_{i} f(x) \delta(x - x_i) dx = \int [f(x) \delta(x - x_1) + f(x) \delta(x - x_2) + \cdots + f(x) \delta(x - x_n)] dx
\]

\[
= f(x_1) + f(x_2) + \cdots + f(x_n) = \sum_{i} f(x_i)
\]

We therefore derive gravitational mass additivity provided that we can motivate the idea that the mass of an object with a complex Dirac delta function is its integral. We can motivate this idea by holding that it is analytic: by definition, the value of the property designated by a Dirac delta is its integral. The substantive part of the proof is the transformation of the fundamental equation into an equation that makes clear that the composite mass density is one whose integral yields mass additivity. This is a contingent result, which depends on the form of the law. For example, if the field were a function of the square of the density then we would not derive mass additivity despite the analytic stipulation. It is therefore a legitimate proof that mass additivity is grounded in the form the fundamental law.
Note that a similar proof for inertial mass additivity suffers two problems. Firstly, one would have to represent acceleration as a density with a Dirac delta function. But arguably, the acceleration of a composite is a vague point value, not a distribution of precise point values. Secondly, the form of the law would require that we multiply the mass density with the acceleration density. But multiplication of Dirac Delta functions is not well defined. Thus, this style of proof is better suited to gravitational mass additivity.

2.7 Charge

In this section I defend an a priori entailment expansion for the following conditional:

(Charge additivity): If [Newtonian microphysical description] & [T] then charge is additive.

Electric charge is similar to gravitational mass. Just as gravitational mass is a property of objects that causes them to experience a force when near other (gravitationally) massive objects, so too is electric charge a property of objects that causes them to experience a force when near other electrically charged objects. Electric charges can be positive or negative. Positively charged objects, such as protons attract negatively charged objects such as electrons, and vice versa. Electric charge is additive: the charge of a composite is the sum of the electric charges of its parts. However, the macroscopic effect of charge additivity is relatively minimal because macroscopic objects are made of atoms and atoms typically have equal numbers of protons and electrons such that their charges cancel out (neutrons also compose atoms but have no charge or are “neutral”).

Due to the nearly identical structure of the gravitation law and the charge law (Coulomb’s law), the required expansion need only mimic the gravitation expansion. Let’s begin with the law governing point charges in pre-relativistic classical physics. The electric field $E(x)$ that a single point charge $q_1$, located at $x_1$ determines, for some position $x$, is as follows:

$$E(x) = \frac{1}{4\pi\epsilon_0} \frac{q_1(x - x_1)}{|x - x_1|^3}$$  \hfill (37)

The law for $N$ charges invokes a composition law:

$$E(x) = \frac{1}{4\pi\epsilon_0} \sum_{i=1}^{N} \left[ \frac{(x - x_i)}{|x - x_i|^3} q_i \right]$$  \hfill (38)

We may now expand (Charge additivity):
(1) **Newtonian microphysical description:**

There are two charges $q_1$ and $q_2$, located in Euclidean space at the same position $x_1$. They determine an electric field $E(x)$ according to the composition law and Coulomb’s law:

$$E(x) = \frac{1}{4\pi\varepsilon_0} \sum_{i=1}^{N} \frac{(x-x_i)}{|x-x_i|^3}q_i$$

That’s all that is fundamental.

(2) **[Composition]** If (1) then there is a composite $C$ composed of $q_1$ and $q_2$ located where $q_1$ and $q_2$ are located (point $x_1$), in virtue of $q_1$ being at $x_1$ and $q_2$ being at $x_1$.

(3) **[Field]** If (1) & (2) then $C$ determines electric field $E(x)$, in virtue of $C$’s parts determining $E(x)$.

(4) **[Overlap Charge]** If (1)-(4) then $C$’s charge-role is being played by $[q1+q2]$ in virtue of $E(x) = \frac{1}{4\pi\varepsilon_0} \frac{(x-x_1)}{|x-x_1|^3}(q_1 + q_2)$. So charge is additive for overlapping composites.

(5) **[Compact Charge]** If (1)-(5) then $C^*$’s charge-role is played by $[q1+q2]$ in virtue of $E(x) = \frac{1}{4\pi\varepsilon_0} \frac{(x-x^*)}{|x-x^*|^3}(q_1 + q_2)$, where $x^*$ is a member of the highly constrained set of admissible idealisations for $C^*$’s position. So charge is additive for compact composites.

(6) **[Arbitrary-Fusion Charge]** If (1)-(6) then $C^{**}$’s charge-role is played by $[q1+q2]$ in virtue of $[q1+q2]$ playing $C$’s charge-role in both overlap and compact situations. So charge is additive for arbitrary fusions.

(7) **[Charge Additivity]** If (1)-(8) then charge is additive.

(8) **Macrophysical description:** charge is additive.

As the expansion conditionals are effectively identical to the expansion conditionals from section 2.5 (the only difference being that we are dealing with charges rather than masses and electric fields rather than gravitational fields), I defer the reader to section 2.5, as well as to the more extensive discussion of the epistemology of mereology in section 2.8. Premises (4)-(6) depend on the analysis of charge. One can characterise the charge-role on the model of the gravitational mass-role, provided that one appeals to the disposition to *repel*, as well as to attract. That is, to play the charge-role is to be...
the most natural measure of the disposition of objects to attract or repel other objects as they come closer. We can also run an argument analogous to the one in section 2.6., which derives charge density additivity.

2.8 Mereology

This section provides a more detailed defence of the apriority of the [Composition] expansion conditionals. Each [Composition] conditional inferred the existence of a composite from the existence of two elementary particles (e.g. two masses or two charges). Their general form is:

[Composition] If [Fundamental microphysical description] and [T] then there is a composite C composed of microphysical o₁ and o₂, that is located where o₁ and o₂ are located (the set of points \( \{x₁, x₂\} \)), in virtue of o₁ being at x₁ and o₂ being at x₂.

I defended the apriority of these conditionals by briefly defending the apriority of unrestricted composition. Unrestricted composition states that any objects compose an object (their composite). In defence of the apriority of unrestricted composition I appealed to the Analysis Based Justification heuristic. Here I develop these arguments. I also show that the a priori entailment expansions, and hence AET, only require a weaker claim than the apriority of unrestricted composition.

Unrestricted composition is an answer to what Van Inwagen (1990) calls "the special composition question": under what circumstances do objects compose wholes? Unrestricted composition is controversial, however. Some philosophers accept restricted composition and argue that objects compose wholes only under certain circumstances. Others accept mereological nihilism and argue that composition never takes place: there are only elementary atomic particles. In my view, restricted composition and mereological nihilism can be ruled out a priori.

Although unrestricted composition is debated in the metaphysics literature, metaphysicians don’t assert or deny it based on any experimental evidence. I sometimes quantify over composites composed of two photons travelling in different directions, to demonstrate that massive composites can sometimes be composed of massless parts. So do many physicists. Now, what experiment could a critic of unrestricted composition do on photons, or on whatever, to show that to the contrary, the two
photons do not in fact compose a composite? It’s not as if technology fails us here, the very idea of performing such an experiment is misplaced.

There is therefore a strong case that this metaphysical debate is a merely verbal dispute. The concept of a composite I intend when stating [Composition], is applied by grouping objects in thought, formulating a name for them such as 'composite C', and applying that name “C exists”. That is its semantics. To be justified in believing that C exists one only needs to be justified in believing in the grouped objects. Possessing the concept comes with the ability to consider certain objects, group them in thought, and name them. This ability grounds the apriority of expansion conditionals with antecedents that quantify over components, and consequents that quantifier over their composites.

There can be no possible counterexample to unrestricted composition because any possible group of objects that might be thought to be a counterexample can easily be grouped, named and quantified over. So deniers of unrestricted composition cannot be using 'composite' to express this concept. In what follows I designate this concept using capital letters and a subscript—COMPOSITE\(_U\).

Metaphysicians who reject unrestricted composition must either take COMPOSITE\(_U\) to be defective, or to be legitimate but not the concept at issue. Many metaphysicians think the application conditions of the concept at issue are neutral on the special composition question. One way of developing this idea is by disambiguating a very thin concept—call it COMPOSITE\(_*\)—that applies to objects just in case they have parts. The debate is then about whether there are any composites\(_*\) (objects with parts) and if so, the situations under which objects are parts of other objects.

But this just pushes the problem from the notion of 'composite' to the notion of 'part'. Thus, the consequent of [Composition] could instead state "there exists an entity that has m\(_1\) and m\(_2\) as parts". Although 'composite' is not mentioned in this formulation of [Composition], the defence of its apriority remains the same. For we apply 'part' to an entity e just in case we can group e together with other entities and apply a name 'C' to the group. We can then say "e is a part of C". So defining 'composite' solely in terms of 'part' does not help. Nor does it help to assert that 'part' expresses a concept just as thin as COMPOSITE\(_*\), or that it expresses a primitive that has no associated application conditions. For this just yields notions of 'part' and 'composite' that are not associated with any rules for their application. Such inapplicable concepts are defective.

53 Hirsch (2002) is an influential defence of this deflationary diagnosis. Hirsch argues that the disputants employ distinct quantifiers, meaning different things by the phrase ‘there exists something’. I believe it more accurate to diagnose the dispute in terms of distinct notions of 'composite’, but my arguments could equally well be phrased either way.

54 Bohn (2009) argues that worlds in which every object is a proper part of some other object (“junky worlds”) are possible counterexamples, because they don’t contain a universal object composed of everything. Watson (2010) responds that such worlds are not possible. Contessa (2012) responds that unrestricted composition does not require the existence of a universal object.
A simple fix—to turn COMPOSITE* into a non-defective concept—is to enrich it so that possessors of it have, in virtue of possessing it, some way of applying it to cases. Using subscripts where U=universal composition, R=restricted composition, and N=nihilism, we then have three concepts, COMPOSITE_U, COMPOSITE_R, AND COMPOSITE_N. The first has very simple application conditions: if there are at least two objects then there is a composite. The second does too, although the restriction may need to be spelled out; perhaps: if there are some non-scattered causally integrated objects then there is a composite. It is not clear to me that the third concept can in fact be spelled out—ultimately I would conclude that it is a defective concept.

It is sometimes said that these application conditions cannot be genuine a priori or analytic truths, simply because they are existence claims. For example, consider Chalmers' concern (I’ve substituted 'composite' for 'sum'):

"Consider ontological theses such as [...] 'If there are two objects, there is a composite of those two objects'. [Some philosophers] hold that theses of this sort are trivially true (or false), and often hold that they are analytically true (or false). But no existence claim is trivially true, and certainly no existence claim is analytically true. It may be analytic or trivial that 'If there is an object composed wholly of those two objects then there is a composite.' But the claim that there is an object with those properties [...] is never trivially or analytically entailed by a sentence that does not make a corresponding existence claim."

(Chalmers, 2009: 77)

The notion of a corresponding existence claim is unclear in this passage. For why is it not the case that 'there are two objects' makes a “corresponding” existence claim to 'there is a composite of those two objects'? We can agree that no existence claim is trivial or analytic. But it is implausible that there are no analytic or a priori conditionals with one existence claim in the antecedent and a different existence claim in the consequent. There are many examples:

'If there exists an unmarried man then there exists a bachelor.'

'If there exists only one 5 foot man and one 6 foot man then there exists a shorter person and a taller person.'

Chalmers would not in principle deny the analyticity of such conditionals, so it must be that their antecedents make existence claims that “correspond” to the existence claims in the consequents.

One way of clarifying ‘correspond’ is in terms of definability: being a bachelor is (at least partly) defined in terms of being an unmarried man therefore existence claims about bachelors and unmarried men “correspond”. But if anything this supports the idea Chalmers is criticising: being a composite is defined in terms of being two (or more) objects, therefore existence claims about composites and
atoms “correspond”. I have defended this definitional claim in detail by explaining the notion so defined, and using the notion in a number of proofs.

Another way of clarifying ‘correspond’ is in terms of *identity*. That is, perhaps the existence claims correspond when the being whose existence is asserted in the consequent is the same being whose existence is asserted in the antecedent.\(^{55}\) Presumably the unmarried man and the bachelor are identical—there is just one man falling under two different descriptions. In that case, perhaps the symmetry is broken if the composite is not identical to its parts. Given the way I have so far characterised \(\text{COMPOSITE}_U\), the concept appears committed to composition as identity. For applying the concept essentially introduces a singular term ‘composite such and such’ for a plurality of objects. But composition as identity is a problematic thesis, as composites appear to have distinct properties from their parts.

Take a composite composed of two particles. One objection to composition as identity is that the composite has the property of being one in number, while the parts have the distinct property of being two in number. This is a bad objection because the composite also has the property of being two in number: it has the property of being two particles. The objection presupposes that it is incompatible to have the property of being two things while simultaneously having the property of being one thing. But it is perfectly compatible: any two things is also one composite. As David Lewis put it:

“[T]he many and the one are the same portion of Reality, and the character of that portion is given once and for all whether we take it as many or take it as one [...] It does matter how you slice it—not to the character of what’s described, of course, but to the form of the description”.

(Lewis 1991:87)

A better objection to composition as identity appeals to the apparent fact that the parts may have the property of being able to survive certain events that the composite cannot. For example the bush will not survive the bushfire, but its elementary components will—after all, they’ve been around since the early universe. Furthermore, the composite may have the property of being possibly made of distinct parts: the bush could have been made of more parts than it is actually made of, if the slug did not eat the leaf (say). But this does not appear to be true of the parts.

In response, we must clarify \(\text{COMPOSITE}_U\) further. So far, I have characterised its application conditions in terms of grouping objects and naming them. But one can also group objects and name them *in virtue of their configuration*. Thus, one might take objects configured as a square, and name them ‘composite square 1’. And then as soon as they lose their square configuration, they are to be no longer deemed ‘composite square 1’ as a matter of stipulation. Here composite object 1 is not

\(^{55}\) This is Chalmers’ preferred reading.
identical to its parts, because its parts can survive disconfiguration. Thus, composition as identity is true for a subset of composites. But insofar as \textsc{composition}_U allows that naming can be constrained by configuration, then \textsc{composition}_U is not committed to composition as identity. Now consider the following conditional:

‘If there are objects arranged square-wise then composite square 1 exists’

It is true that ‘there are objects’ and ‘composite square 1 exists’ are not “corresponding” existence claims. That’s because composite square 1 is not identical to the objects. However, the consequent and the full antecedent are corresponding existence claims. For composite square 1 is identical to the objects arranged square-wise. The symmetry between the composition conditional and the bachelor conditional is therefore not broken.

One should also allow application conditions for \textsc{composite}_U so that it applies vaguely to sets of admissible precisifications. Thus, one could apply \textsc{composite}_U by considering some objects, grouping various precise subsets of those objects in thought and naming them ‘subset 1’, ‘subset 2’ etc. and then introducing the name ‘vague composite 1’ so that it applies vaguely to the subsets. The subsets are admissible precisifications of ‘vague composite 1’, although ‘vague composite 1’ does not determinately apply to any one of the subsets more than it does to any other. Now consider the following conditional:

‘If there are objects arranged thus and so then vague composite 1 exists’

Here it is harder to see how the existence claims in the antecedent and consequent correspond. However, there is enough correspondence to retain the symmetry: the admissible precisifications of vague composite 1 are each identical to certain of the objects arranged thus and so.

I thus conclude that Chalmers’ objection does not work: unrestricted composition can be defended through conceptual clarification alone, and is therefore a priori. With these application conditions for \textsc{composite}_U clarified, I hope it is clear that all composites are a priori deducible. Note that the masses of vague composites will be as vague as the composites themselves: their admissible precisifications will be the masses of the admissible precisifications of the vague composites.

I think the most promising objection concedes that \textsc{composite}* requires further application conditions built into it, but argues that they can be specified in a less committal way. Here one might appeal to a naturalness constraint. Thus, consider the concept \textsc{composite}_{JC} where JC means “joint carver” (using "joint carver" rather than "naturalness" saves us from using subscript N again). The application conditions for this concept are spelled out as follows: if there are some objects such that the fundamental structure of reality groups those objects by some natural parthood-like relation, then there is a composite. Thus, perhaps non-scattered causal integration will turn out to be the most
natural manner in which elementary particles are grouped in nature, but perhaps some other natural grouping mechanism will surface in future empirical enquiry. The idea is that whatever the natural grouping mechanism is, it is not knowable a priori, and so unrestricted composition is not a priori, and is potentially a posteriori false. This is essentially Sider’s (2009) response:

Let us henceforth conduct our debate using ‘∃’. We hereby stipulate that ‘∃’ is to express an austere relative of the ordinary English notion of existence. We hereby stipulate that although the meaning of ‘∃’ is to obey the core inferential role of English quantifiers, ordinary, casual use of disputed sentences involving ‘there exists’ (such as ‘Tables exist’) are not to affect at all what we mean by ‘∃’. We hereby stipulate that if there is a highly natural meaning that satisfies these constraints, then that is what we mean by ‘∃’. Perhaps the resulting ‘∃’ has no synonym in English. Fine—we hereby dub our new language Ontologese.

(Sider 2009: 415-6 [my italics])

Note that ‘Tables exist’ is only disputed because ‘If there exists particles arranged table-wise, then there are tables’ is disputed by philosophers who accept the antecedent. This is why it is better to frame the debate in terms of the notion of ‘composite’ rather than ‘exists’: mere existence is not at issue, what is at issue is whether composites exist given the existence of mereological simples. The objection can then be understood as postulating a notion of ‘composite’ with the most natural meaning, which also satisfies basic constraints on the ordinary notion of ‘composite’.

In reply: one way to determine a priori whether the extensions of COMPOSITE\textsubscript{JC} and COMPOSITE\textsubscript{U} ever come apart is to consider (as actual) as many worlds as possible to see if any sense can be made, in any world, of “the most natural manner in which elementary particles are grouped in nature”. This is an a priori exercise: imagine an ideal reasoner who possesses all possible concepts--she can construct all possibilities in thought, and can therefore determine whether any possibility suggests a natural restricted grouping mechanism. She can do this without consulting which world is actual. So the modal equivalence of COMPOSITE\textsubscript{JC} and COMPOSITION\textsubscript{U} can be determined a priori. The question is whether we can determine it a priori with our limited conceptual repertoire.

We can determine the equivalence of COMPOSITE\textsubscript{JC} and COMPOSITION\textsubscript{U} because we can generalise from paradigmatic possible worlds. In particular, none of the worlds described by physical theories we know of group parts into wholes in some restricted way. As already argued, nothing in relativisitic physics prevents one from quantifying over the composite of two photons travelling in different directions and it is sometimes quite useful to do so. The problem with any imposed restriction, such as the non-scattered causal integration constraint, is that they are too subjective and too dependent upon human needs and interests and the general ways our brains prefer to efficiently process information. The fundamental elements of clearly conceivable close possible worlds (e.g.
classical worlds), do not group parts into wholes in any preferred manner, nature simply presents the fundamental structure of reality and leaves the grouping of its fundamental elements up to us and our multi-vari ous interests. It is a priori reflections like this that a priori justifies one in believing unrestricted composition and hence all of the instances of [Composition] that we have discussed.

While the apriority of unrestricted composition is useful in the exposition of AET, AET does require it. What's required is just the apriority of conditionals of the form: If [fundamental description] then [the correct composition principle]. So even if COMPOSITE_{3C} picks out a restricted set of particle groupings, AET is defensible if the motivation for the restriction is either a priori deducible from the fundamental truths, or explicitly stated by the fundamental truths.

AET is also compatible with restricted composition. An advocate of restricted composition may reject [Composition] based on their intuition that two point particles are “not enough” to compose a composite. But to infer all the composites_{R} one just needs to infer all the composites_{S} that meet the relevant restriction. So if the restriction is 'high causal integration among parts' then one just needs to find the composites_{S} with high causal integration to deduce the composites_{R}.

Even a version of nihilism is consistent with AET and suitably formulated versions of our expansions. One version of nihilism states that reality can be distinguished into distinct levels of property exemplification. There is the fundamental microphysical level consisting in atomic particles and their physical properties. There are then higher levels of property exemplification. However, individual objects do not exemplify these properties, only pluralities of atomic particles do. If this view is correct then we can simply run the expansions without the [Composition] expansion conditionals. Thus, rather than moving from some objects to their composite then to the composite's position, acceleration, and force, we instead move from some objects to their collective position, acceleration, and force.

A more extreme version of nihilism denies that reality can be distinguished into levels in any sense: not only are there no objects other than atomic particles, there are no properties other than those instantiated by individual atomic particles. Such a view is consistent with AET because AET only states that nonfundamental truths, if any, follow a priori from fundamental truths. Extreme nihilism denies that there are any nonfundamental truths. However, AET is trivialised by this version of nihilism. Fortunately, extreme nihilism is very implausible: we only posit atoms in the first place to explain macroscopic objects and their behaviour. Macroscopic objects are a datum to be explained not a posit to be eliminated.

The consequent of [Composition] states the existence of all the composites and their locations. Accordingly, one might object that we haven’t ruled out a priori the existence of other objects that could disrupt the expansion. How do we rule out the existence of a non-fundamental simple m_{3}, which
is “grounded” in \( m_1 \) and \( m_2 \) without being composed of them? And why is \( C \) the only composite? Why isn’t there also \( C^\wedge \), which is composed of \( m_1 \) and \( m_2 \) without being located where \( m_1 \) and \( m_2 \) are?

The \( C^\wedge \) hypothesis is inconceivable: an object spatially distinct from \( m_1 \) and \( m_2 \) cannot be inferred by grouping \( m_1 \) and \( m_2 \) in thought and naming them. For the name goes wherever \( m_1 \) and \( m_2 \) go. So insofar as \( \text{COMPOSITION}_U \) is expressed by ‘composition’ in [Composition], the \( C^\wedge \) hypothesis can be ruled out a priori. It is not clear whether the other disambiguations of ‘composition’ allow us to make sense of this hypothesis. The \( C^\wedge \) hypothesis formulated in terms of \( \text{COMPOSITION}_R \) can also be ruled out a priori. Truths invoking \( \text{COMPOSITION}_R \) can be deduced from microphysics by first inferring all the composites \( \text{U} \) and then inferring which of those composites \( \text{S}_U \) are sufficiently causally integrated (or whatever the restriction is). All such composites \( \text{S}_R \) will be located where their parts are located in virtue of all the composites \( \text{U} \), being located where their parts are located.

The \( m_3 \) hypothesis is also inconceivable but for different reasons. The hypothesis is that there exists a non-fundamental mereological simple \( m_3 \) that is grounded in \( m_1 \) and \( m_2 \). This hypothesis can be ruled out by appeal to our concepts of grounding and fundamentality. Our concept of grounding is tied to our concept of non-causal explanation (section 1.1). Because \( m_3 \)’s existence cannot be explained in terms of \( m_1 \) and \( m_2 \) we are justified in ruling out the hypothesis. One does not explain \( m_3 \)’s existence by simply stating that \( m_3 \) obtains in virtue of \( m_1 \) and \( m_2 \), just as one does not (and cannot) explain an economy’s existence by simply stating that the economy obtains in virtue of \( m_1 \) and \( m_2 \). As discussed in sections 2.3 and 3.1, real scientific reductive explanations explain why the explanandum exists in virtue of explanans by showing that features of the explanans play the conceptual role that the explanandum plays. And just as no features of \( m_1 \) and \( m_2 \) alone can conceivably play the role played by an entire economy, no features of \( m_1 \) and \( m_2 \) alone can conceivably play the \( m_3 \) role. Such considerations enable one to rule the \( m_3 \) hypothesis out a priori.

I thus conclude that the apriority of the [Composition] expansion conditionals is well founded.

2.9 Beyond Classical Physics

Where does the argument for the apriority of (Inertial mass additivity) go wrong in the relativistic context? The difference comes at premise (5) [Uniac mass]. This can be demonstrated most simply in a highly restricted situation where the relativistic equation is similar to the Newtonian equation. In this simple situation there are two (or more) particles with identical accelerations perpendicular to the

---

56 Similarly Jackson (2003b: 159) argues “someone who thinks that three electrons might make up a shopping centre does not have our concept of a shopping centre”. See also Jackson (2006a).
direction of their velocities. (Imagine a swarm of electrons with velocities perpendicular to a uniform magnetic field.) This situation simplifies things because it allows us to use three vectors, without having to make explicit the equation for when force is parallel to velocity. The superposition and composition of forces hold in relativity, so we may derive the additivity of forces to give the force on the composite C:

$$F_c = \sum_{i=1}^{N} a_i \gamma_i m_i$$  \hspace{1cm} (39)

Because component accelerations are identical, a simple algebraic transformation enables us to take the acceleration term out of the scope of the sum:

$$F_c = a_1 \sum_{i=1}^{N} \gamma_i m_i$$  \hspace{1cm} (40)

The subscript ‘i’ has been replaced with 1, designating component particle 1. Since the accelerations of all the parts are identical to $a_1$ we can appeal to restricted acceleration composition (from section 2.4) to derive the following result, which relates the force and acceleration of the composite, to the masses (and gamma factors) of its parts:

$$F_c = a_c \sum_{i=1}^{N} \gamma_i m_i$$  \hspace{1cm} (41)

We now appeal to our a priori definition of mass:

**Definition of Inertial mass:** The mass of an object is a measure of the object’s most natural disposition to resist changes in motion given applied forces.

If we reflect on equation (41), we see that the sum of the products of the component masses and gamma factors is the coefficient relating composite force to composite acceleration. So (41) entails that the extent to which C changes motion in response to a force is proportional to the sum of the products of the component masses and gamma factors. Hence, we can infer that the composite’s disposition to resist changes in motion given applied forces is the sum of the products of the component masses and gamma factors. But this is not a particularly natural property due to the dependency of $\gamma$ on an arbitrary frame of reference. And here is where Newtonian physics and relativistic physics come apart.

The value $\gamma_i m_i$ is the relativistic mass of particle i. So our procedure correctly derives the additivity of relativistic mass. But what about mass? Here we find the C’s frame invariant properties and define its most natural properties accordingly. For this we consider the rest frame, the frame in which gamma...
equals one. So the most natural property responsible for a particle’s inertia is the property responsible for the particle’s inertia in its rest frame. In this frame total momentum of components is zero. Typically, components move relative to the composite rest frame, and thereby alter the inertia of the composite (section 2.2). So our procedure gives the right results for mass in a Newtonian world and mass in a relativistic world.

The above three-force equations are non-standard representations. Let us then consider the standard explanation of mass non-additivity in standard formalism, and ask whether it can be converted into an a priori entailment expansion. The following is taken from Gabovich and Gabovich’s (2007) paper, entitled *How to explain the non-zero mass of electromagnetic radiation consisting of zero-mass photons*:

“The non-additivity of masses can be easily obtained from the *additivity* of conserved properties: energies and momenta for particles in question. Namely,

\[ E_{12} = E_1 + E_2 \] (42)

And

\[ p_{12} = p_1 + p_2 \] (43)

At the same time a dispersion law of the free elementary particle in the special theory of relativity has a more sophisticated form than in Newtonian mechanics [..]:

\[ E = c\sqrt{p^2 + m^2c^2} \] (44)

From equations (42), (43) and (44) one obtains the mass \( m_{12} \) of the composite object:

\[ m_{12} = \sqrt{\frac{E_{12}^2}{c^4} - \frac{p_{12}^2}{c^2}} = \sqrt{\frac{(E_1 + E_2)^2}{c^4} - \frac{(p_1 + p_2)^2}{c^2}} \] (45)

One can easily ascertain that \( m_{12} \neq m_1 + m_2 \), contrary to what is characteristic of the non-relativistic case.”

(Gabovich and Gabovich 2007: 650-1)

Equation (44) is fundamental, while equations (42) and (43) are additivity principles that will need to be derived from fundamentals, before (45) can be derived. Then, when (45) is derived, it will have to be shown that \( m_{12} \) plays the mass role. If this can be achieved, then the principle relating composite mass to fundamental properties of elementary particles is a priori deducible from fundamental relativistic microphysics.
Now, one often finds (e.g. in Lange 2002: 221) the following two definitions: (i) the net change in a body's momentum equals the accumulation of force felt by the body over a given span of time and (ii) the net change in a body's energy is equal to the accumulated force felt by a body as it moves across a certain distance. If these definitions are a priori then one can use them to derive energy and momentum additivity from force additivity. An a priori entailment would rigorously construct each step one at a time, and show how each step depends on certain previous steps. I leave the details to future research.

I have shown that a priori inferences from the fundamental descriptions of classical worlds enable one to infer the non-fundamental truths of such worlds. It is natural to wonder why this works and whether we should always expect this from fundamental theories. This is a difficult question. There are hypotheses about the nature of fundamentality and grounding that entail AET. For example, the hypothesis that fundamentality is a property only of propositions, and grounding is a relation only between propositions. On this hypothesis, there are no "levels of reality" there is only reality, which can be described in terms of different "levels of proposition". Here, propositions (or families of propositions) can be distinguished into "levels", where the fundamental level describes reality to the highest degree of detail in the most minimal and precise vocabulary. All other (families of) propositions are more coarse grained descriptions of reality, but are nonetheless true of reality because the coarse grained information they contain can be found within the fine-grained descriptions (e.g. as patterns). Such a view, I believe, is supported by the fact that AET appears to hold for several physical theories. And insofar as such a view is plausible, then we should expect AET to be true of any successful fundamental physical theory. I say more about this theory in section 3.4. Next, I consider the relationship between AET and explanation.
Evaluating the Reduction Entailment Thesis

3.1 The Objection from Real Reductions

Recall the Reduction Entailment thesis:

Reduction Entailment thesis (RET): For any proposed reductive explanation, if the explanans does not a priori entail the explanandum, then the explanation is either incomplete or defective.

Reductive explanations answer grounding questions by illuminating why features of the explanandum exist, in a world containing the explanans. RET implies that a priori entailment is a necessary condition on successful reductive explanation.

One might naturally wonder why a priori entailment should be connected with explanation at all. The thought behind RET is that if a reductive explanation exhibits a priori entailment, in the sense that the explanans a priori entails the explanandum, then this positively contributes to the explanation’s ability to illuminate why the explanandum obtains in virtue of the explanans. To get a sense of this kind of explanatory contribution, consider the following toy example: imagine explaining to someone why Smith is taller than Jones by appeal to the fact that Smith is six foot tall and Jones is five foot tall. Here, the explanans a priori entails the explanandum. Notice that the explanans does not logically entail the explanandum—the inference from explanans to explanandum is not truth-preserving in virtue of logical form alone. Not all a priori entailments are logical entailments and this point is crucial for when we consider inhomogeneous reductions (section 3.3). Still, the explanation clearly exhibits a priori entailment. After all, the corresponding conditional that takes the form ‘If [explanans] then [explanandum]’ exhibits the usual features we associate with a priori truths. Understanding it puts one in a position to be justified in believing it; empirically testing it seems redundant; and its negation is inconceivable. Arguably, these features contribute to the success of the explanation. Accordingly, a reductive explanation that fails to exhibit a priori entailment is missing a positive epistemic feature that it could otherwise have. The idea behind RET is that if a proposed reduction is missing this feature, then the explanation is either incomplete or defective.
If an explanation exhibits a priori entailment, then the material conditional with the explanandum as its consequent and the explanans as its antecedent, exhibits the apriority symptoms. Let’s consider why this is an explanatory virtue for reductive explanations. If one can know the conditional simply by understanding it, then one is able to see a tight connection between explanandum and explanans. If testing the conditional is redundant, then no coherent question of the form ‘what if empirical method such and such enabled us to discover that the explanans can obtain without the explanandum?’ can arise. Similarly, if one cannot conceive the negation of the conditional, then no coherent question of the form ‘what if we are in such and such a world in which the explanans obtains without the explanandum?’ can arise. Thus, exhibiting a priori entailment strongly suggests that the explanation answers all meaningful questions about why the explanandum exists in a world containing the explanans.

But perhaps a priori entailment is simply too much to ask for. One could demonstrate this if one could find examples of intuitively successful reductive explanations that do not exhibit a priori entailment. Indeed, an influential objection to RET is that, on the face of it, no scientific reductive explanation exhibits a priori entailment from explanans to explanandum, including uncontroversially successful reductions. This brings us to the objection from real reductions:

**Objection from Real Reductions:**

1. If successful scientific reductive explanations do not exhibit a priori entailment, then RET is false.
2. Successful scientific reductive explanations do not exhibit a priori entailment.
3. Hence, RET is false.

Whatever the initial plausibility of the argument, it relies on a crucial ambiguity in the key notion of exhibiting a priori entailment. In particular we must distinguish between directly exhibiting and indirectly exhibiting a priori entailment. If we read the argument in terms of ‘directly exhibiting’ then although premise (2) is plausible, premise (1) is implausible. Meanwhile, if we read the argument in terms of ‘indirectly exhibiting’ then although premise (1) is plausible, premise (2) is implausible.

An explanation, as presented on paper or in speech, directly exhibits a priori entailment if and only if the explanans as it is stated a priori entails the explanandum as it is stated. It is implausible that real explanations directly exhibit a priori entailment. If they did they would be much more complicated than they are. However, RET does not require scientific reductions to directly exhibit a priori entailments—that’s too much to ask. Rather, RET only requires that reductions indirectly exhibit a priori entailment. An explanation indirectly exhibits a priori entailment if and only if the explanation gets you half way there: if it sets up the beginnings of an a priori entailment so that understanding the explanation gives one good reason to believe that an explanation that directly exhibits a priori entailment is possible. RET deems such explanations to be incomplete rather than defective. If an
explanation gives no suggestion as to how a priori entailment could obtain—i.e. does not even indirectly exhibit a priori entailment—then RET deems is defective.

The RET advocate must treat all successful reductive explanations as incomplete as stated. But what reason do we have to consider intuitively successful reductions to be incomplete? To respond to this challenge, RET advocates owe us an account of incompleteness. But, ‘incomplete’ can’t simply mean ‘does not exhibit a priori entailment’—for this just presupposes that explanation requires a priori entailment. Thus, RET advocates require a non-circular account of the completeness of an explanation that enables a response to the pressing challenge from real reductions. Meeting this challenge is the task of this section.

The account I defend states that uncontroversially successful reductions which do not directly exhibit a priori entailment are incomplete in the sense that they exhibit subtle explanatory gaps (that is, their explanans leave open subtle yet explanatorily relevant questions about the explanandum). Furthermore, were we to enrich the explanans so as to close those explanatory gaps, the explanation would directly exhibit a priori entailment.

Let’s illustrate the objection from real reductions with the reductive explanation of Newtonian inertial mass additivity. This is a good starting point because it is a relatively clear example of a successful reductive explanation that does not directly exhibit a priori entailment. Furthermore, in the previous chapter we ‘completed’ the explanation thereby demonstrating that the explanation indirectly exhibited a priori entailment. This means that here, we can just provide a schematic completion of the explanation, in order to defend my response to the objection.

There are, however, several limitations with the example. It is an example of a homogeneous reduction: the vocabulary of the explanandum is contained in the explanans (specifically, the key term ‘mass’). So it’s not clear that we can generalise from such a case to cases of inhomogeneous reductions, where the explanandum introduces new vocabulary. However, I will argue that RET need not distinguish between homogeneous and inhomogeneous reductions—they can be treated on a par. I then discuss inhomogeneous reductions in more detail in section 3.3.

Let’s begin by distinguishing the explanandum, the explanans, and the explanatory premises appealed to, which are intended to show why the explanans holds in virtue of the explanandum. Note that this is a simplified version of the Kibble and Berkshire explanation from section 2.3.

**Explanandum:** mass is additive: the mass of a composite is the sum of the masses of its atomic component parts.

**Explanans:** three particles \( p_1 \) and \( p_2 \) and \( p_3 \); their physical states (acceleration, mass); force laws applied to the two particles: \( F=MA \) (the Force on a particle is equivalent to its Mass
times its Acceleration), and, \( F_i = \sum_{j=1}^{N} F_{ij} \) (the force on particle \( i \) is the sum of the forces on \( i \) due to the objects indexed by \( j \)).

**Explanatory premises:**

Composite \( C_{12} \) exists in virtue of \( p_1 \) and \( p_2 \).

\( C_{12} \) has force \( F \) in virtue of \( p_1 \) and \( p_2 \) having force \( F \).

\( C_{12} \)'s acceleration is identical to the accelerations of its parts \( (a_c = a_1 = a_2) \).

Using \( C_{12} \)'s force and acceleration we can solve for mass and infer mass additivity.

This explanation does not directly exhibit a priori entailment for two reasons:

(i) Explanans only appeals to a two-particle system and does not a priori entail relevant facts about more complex systems that are within the scope of the explanandum.

(ii) Explanans only appeals to composites whose parts have identical accelerations and does not a priori entail facts about composites within the scope of the explanandum whose parts have distinct accelerations.

Kibble and Berkshire treat this explanation as unproblematic (just as Lindsay (1961) does with his variant). The explanation is therefore thought to be successful by certain authorities on the issue. There is thus a prima facie reason to treat the explanation as successful. However, the explanation does not exhibit a priori entailment for the two reasons given. So it looks like a good example to illustrate the objection from real reductions.

However, (i) and (ii) arguably constitute *explanatory gaps*. An explanatory gap is an explanatorily relevant question about the explanandum that is left open by the explanans. The explanation does not obviously extend to the three-particle system, or to more complex systems. For perhaps mass additivity fails for a composite composed of four particles? The explanans appears to leave this question open. Secondly, the explanation does not obviously extend to composites whose parts do not have identical accelerations. How do we know that mass additivity does not fail for such composites? The explanation appears to leave this question open too. Both of these look like questions that the explanation as stated, leaves open. Here we get a sense of explanatory incompleteness without presupposing RET.

A natural defence of the explanation states that we know from experience that composites, whose parts have non-identical accelerations, typically do not differ in any relevant way to composites whose parts have identical accelerations. A similar response could be given for the particle number problem. This a posteriori background knowledge is relatively obvious. And presumably this is part of why the
scientific explanation didn’t explicitly deal with the more complex cases. This brings out both the sense in which the explanation as it stands is successful and also why it does not exhibit a priori entailment. It does not exhibit a priori entailment because the inference from facts about 2-particle systems with non-varying accelerations to many-particle systems with varying accelerations is not a priori. Such an inference requires further a posteriori principles relating the simple system described in the explanans and the more complex systems that are within the scope of the explanandum.

Still, there is something inadequate about appealing to these obvious a posteriori principles to close the acceleration and particle number gaps. For if this was a legitimate move, then we could have just appealed to this style of explanation from the start: forget the standard scientific explanation altogether, we know from experience that when we chop objects in half they typically resist acceleration to half the degree that their composite did, given the same applied force; and we can conclude that mass is probably additive simply from these paradigmatic cases. But this sort of explanation would be considered inadequate in the physics community. Recall that Kibble and Berkshire were determined to show that “one consequence of our basic laws is the additive nature of mass” (2004: 12). This suggests that derivations from fundamentals are important to scientific explanation, at least in this case. So isn’t there value in making sure that it is also a consequence of the basic laws that all relevant systems exhibit mass additivity? Won’t this give us the full reductive explanation of mass additivity?

Let us then enrich the explanans, so as to close these explanatory gaps. In particular, let’s explicitly apply the laws to arbitrarily many particles with arbitrarily distinct accelerations, as we did in section 2.4. The underlined aspects represent the explanatory enrichments:

**Enriched explanans:** uniac situation involving arbitrarily many particles $p_1, p_2, ..., p_n$; their physical states (acceleration, mass) in the actual and nomologically possible uniac (C), near-uniac (C*), and arbitrary-fusion (C**) scenarios; force laws: $F=MA$ (the Force on a particle is equivalent to its Mass times its Acceleration), and the superposition principle for interactions ($F_i = \sum_{j=1}^{N} F_{ij}$) (the force on object $i$ is equivalent to the sum of each of the individual forces $F_{ij}$, where $F_{ij}$ stands for the force on object $i$ due to the object indexed by $j$ ($j=1,2,...,N$)).

Notice that if we frame the explanatory premises in terms of conditionals, then we have the first five steps of the a priori entailment expansion discussed earlier, which expands the conditional ‘If [enriched explanans] then mass is additive’:

1. **[Enriched explanans]**
2. **[Composition]** If (1) then there is a composite C composed of $p_1, p_2, ..., p_n$.
3. **[Force]** If (1) & (2) then there is a force F on C, in virtue of the force F on C's parts.
(4) **[Acceleration]** If (1) & (2) then C has acceleration A, in virtue of its parts each having acceleration A.

(5) **[Uniak Mass]** If (1)-(4) then C’s mass-role is being played by \([m_1+m_2+...+m_n]\) in virtue of \(F=A[m_1+m_2+...+m_n]\). So mass is additive for uniak composites.

(6) **[Near-Uniak Mass]** If (1)-(5) then C*’s mass-role is played by \([m_1+m_2+...+m_n]\) in virtue of \(F=A^*[m_1+m_2+...+m_n]\), where \(A^*\) is an admissible precisification for C*’s acceleration. So mass is additive for near-uniak composites.

(7) **[Arbitrary-Fusion Mass]** If (1)-(6) then C**’s mass-role is played by \([m_1+m_2+...+m_n]\) in virtue of \([m_1+m_2+...+m_n]\) playing C’s mass-role and C*’s mass-role. So mass is additive for arbitrary fusions.

(8) **[Mass Additivity]** If (1)-(7) then mass is additive.

(9) **Macrophysical description**: mass is additive.

Premises (2)-(5), which correspond to the explanatory premises above, as well as (6)-(8), are a priori, as argued in section 2.4. Hence, the target conditional is a priori. Hence, the enriched explanation that closes the explanatory gaps, directly exhibits a priori entailment.

Our case study illustrates a way in which RET advocates can respond to the objection from real reductions: insofar as real reductions don’t exhibit a priori entailment, they exhibit explanatory gaps such that closing them yields an explanation that does exhibit a priori entailment.

Our case study also suggests something of an algorithm for taking successful scientific reductive explanations, and checking whether they confirm or disconfirm RET. Firstly, take a reductive explanation as it has been formulated in a scientific context, one that is widely thought to be a successful explanation.\(^{57}\) Secondly, identify any (usually simplification-induced) explanatory gaps. Thirdly, close these explanatory gaps by enriching the explanans (either by transforming aspects of it, as we did with Newton’s equations, or by adding to it, as with the addition of arbitrarily many Newtonian particles). Fourthly, determine whether the explanandum can be reached from the new explanans by a priori reasoning. If so, RET will receive some confirmation, if not, we have evidence against RET. Whether there are any real reductions that undermine RET in this sense is an open question worth looking into for future research.

Let’s now look to a real instance of an objection from real reductions. In his critical analysis of RET, Marras (2005: 335) argues that RET is “a requirement that no reductive explanation in science should

---

\(^{57}\) There is no reason why these need to be reductions of entire theories to other theories. Arguably, there are few successful examples of such reductions (Silberstein 2002: 94). Reductions of properties, such as being solid or having mass, are sufficient.
be expected to satisfy”.58 The argument is based on his preferred theory of reduction (which involves a controversial interpretation of Nagel’s (1961) classic account of reduction) and how it treats bridge principles. Bridge principles relate properties in the explanans to properties in the explanandum, by identity, or realization, or constitution etc. According to the theory, bridge principles are not a priori derived, rather, they are inductively inferred from a comparison of higher-level explanandum facts and lower-level explanans facts.

To apply the theory to our case study, we have facts stated in point particle mechanics (lower-level facts) and we have facts stated in terms of the mechanics of composites, such as composite mass facts (higher-level facts). From the lower-level facts we derive some other facts (such as Kibble and Berkshire’s algebraic transformation in equation 19). If these other facts are what Marras calls “images” of the higher-level facts, then we have reduced the higher-level facts to the lower level-facts. We then infer bridge principles from reductions like these. However, it is an inductive inference, often involving more than one reduction. Thus, from our scientific explanation in section 2, we might hesitantly postulate a bridge principle holding between the mass of the whole and the mass of the parts (mass additivity). But we might not have a high credence in the postulate until other explanatorily relevant reductions involving mass are given (such as ones involving systems with more than three particles). But even then, we are only given further inductive support for a necessary a posteriori postulate that is inevitably “underdetermined by the evidence” (2005: 351).

Such a brief sketch of Marras’s positive theory of reduction cannot do full justice to it. But I do not wish to criticise it. For perhaps it adequately captures scientific practice, at least regarding reductive explanation and the introduction of bridge principles. The point I want to make is that even if the account is true, it cannot be used to make this objection:

“This classic account of how bridge laws are ‘inductively derived’ from a comparison of two levels of facts [...] flatly contradicts [the] claim that bridge laws are ‘logically supervenient on the low-level facts’, or that they are ‘entailed a priori’ by the latter”.

(Marras 2005: 342, n14)

We should now be in a position to see why there is no contradiction here. It might be that the classic account of how bridge principles are derived contradicts the claim that bridge principles are in practice derived a priori (from lower-level truths); but it does not contradict the claim that bridge principles can in principle be derived a priori (from lower-level truths). Nor does it contradict the claim that in principle derivability is a necessary constraint on successful reductive explanation, for as I have argued, real reductions need not explicitly demonstrate the existence of the underlying a priori

58 Marras (2005: 336, n3) explicitly focuses on the thesis that “[R]eductive explanation of a phenomenon in terms of the [micro]physical requires an a priori implication from the [micro]physical facts to the relevant high-level facts”.
entailments. Still, the a priori entailments need to be there so as to ensure that the subtle explanatory gaps exhibited by real reductions remain innocuous.\textsuperscript{59}

At one point Marras briefly notes that an a priori derivation between truths at different levels could not (in principle!) occur because the truths “are in different theoretical vocabularies and hold in conceptually distinct domains” (2005: 349). This is a common concern that philosophers have with RET: perhaps RET applies to homogeneous reductions, but it surely doesn’t apply to inhomogeneous reductions, whose explananda introduce new vocabulary. And the concern might be used as a reason to think that the example I have appealed to does not take us very far. After all, the reduction of composite mass to basic laws does not bridge distinct vocabularies because the laws contain the word ‘mass’.

In response, RET need not distinguish between the two types of reduction. To see why RET should have no further difficulties accounting for inhomogeneous reductions over and above whatever difficulties it might have in accounting for homogeneous reductions, one must appreciate how much work the analysis of composite position above can do. We can illustrate this with the a priori entailment expansion discussed in section 1.5, which I briefly sketch here. Imagine a two dimensional surface with hundreds of dots distributed over it in certain patterns so that the dots compose various shapes. Imagine this surface is described in rich detail, but only in terms of dots and their Cartesian coordinates on the surface. In particular, the description makes no mention of shapes. The question, then, is whether we can a priori infer the shape-way-things-are from the dot-way-things-are. Arguably, we can. If a series of dots compose a line, then surely we can deduce this simply from the dot-description, provided that (i) we understand the dot-description (ii) we have the cognitive faculties to be able to group objects (dots) in thought so that we may apply a name to their composite (e.g. “line 1”) and (iii) we understand what’s meant by a line. And if we can a priori infer the existence of four such lines of roughly equal length, whose ends are touching, then surely we can deduce the existence of a square. But ‘square’ is a term not contained in the dot-description. So in itself a priori entailment has no problems bridging vocabularies. In section 3.3 I show how this reasoning extends to more physically realistic scenarios.

If we assume that RET is confirmed by successful cases of reductive explanation, then the picture that emerges is one of a priori entailment explanations as ultimate reductive explanations that scientific reductive explanations approximate to, at least insofar as the scientific reductive explanations aim to close explanatory gaps. In that case, at least three important consequences for reductive explanation

\textsuperscript{59} Marras considers my objection: “It might be objected that although we cannot in practice establish inter-level identities a priori, or deduce higher-level truths from lower-level ones, we could do so in principle” (350). Marras responds by pointing out that: “the burden of the foregoing discussion of the Nagelian model was to show that bridge laws are [...] introduced as theoretical hypotheses, and thus as truths of necessity ‘underdetermined by the evidence’” (350-351). But this just restates the view that bridge principles are not introduced by deduction from lower-level physical facts, which as I have argued, is beside the point.
emerge. The first simply concerns understanding the nature of reductive explanation itself. The second and third consequences concern the actual scientific practice of giving reductive explanations. The second consequence concerns ways of improving explanations—at least along one dimension by which we evaluate explanations as being successful. For example, one might want to increase the extent to which one’s reductive explanation reflects metaphysical structure by adding more information to the explanans. If so, what is the relevant information to be added? A potential answer is: information that brings the scientific explanation closer to an a priori entailment explanation. This is because in doing so, one eliminates potential questions about the explanandum, which may have been otherwise left open. Improving the explanation along this dimension will conflict with another dimension along which we evaluate explanations: simplicity. The extent to which one ought to make the trade-off will be sensitive to features of the explanatory context.

The third consequence concerns ways of criticising proposed explanations. One way of criticising a reductive explanation would be to argue that it is unclear how removing its simplification-induced explanatory gaps would yield an a priori entailment explanation. Furthermore, one could argue that a property is irreducible, and perhaps, therefore, fundamental, if one could argue that no reductive explanation of that property in physical terms can be transformed into an a priori entailment explanation through explanatory gap closure. This is one way of making sense of arguments for mind-body dualism in contemporary philosophy of mind: whatever physical explanation we give for consciousness, no amount of simplification-removal will yield an a priori entailment of consciousness. If so, the explanation will not close the primary explanatory gap. I will discuss the reduction of consciousness in more detail in section 3.4. For now, I look to an objection from real reductions that appeals to a reductive explanation that exhibits a rather peculiar explanatory gap.

### 3.2 Reduction of Conductivity

The purpose of this section is to respond to a particular instance of the objection from real reductions, which appeals to an explanation whose explanatory gap, is not entirely obvious. Sometimes, in physics, one explains a macrophysical equation in terms of a microphysical equation, using perturbation theory. According to Max Kistler, the perturbation reductive explanation of the thermal conductivity of metal, refutes the Reduction Entailment Thesis:

“[The example shows that] reduction, i.e. the representation of the solution of the macroequation as an approximation to the solution of the microequation, can only be achieved by using constraints which are essentially inter-level. The lesson [...] should be, it
seems to me, that we cannot construct the solution of the macroequation on the basis of the knowledge of the solution to the microequation alone. Reduction cannot, as is sometimes claimed, be achieved by a priori bottom-up derivation. Even in such an apparently metaphysically innocent case as the thermal conductivity in a metal bar, the solution to the macroscopic equation for heat conservation cannot be derived a priori on the basis of knowledge of the solution to the microequation (for heat conservation) alone. Some input from above (in other words, from knowledge of the macro-level) is needed to constrain the derivation.”

(Kistler 2006: 350)

Indeed, Kistler goes so far as to say that the example also refutes the A Priori Entailment thesis:

“This is an important result. It allows to refute such claims as Chalmers and Jackson's (2001) that the information in a complete microphysical description of the world suffices to deduce all macroscopic truths [...] this is not even true for macro-physical facts.”

(Kistler 2006: 351)

The example that Kistler is referring to is discussed in detail in Rueger (2001, 2006) and in Rueger and McGivern (2010). I will argue that the example is in fact quite problematic, and that it cannot be used to support Kistler's claims.

The example involves the description of a (one-dimensional) metal rod of some length \( L \), where \( T(x) \) is the temperature at position \( x \) and \( K(x) \) is the rod's thermal conductivity at position \( x \):

\[
\frac{d}{dx} \left[ K(x) \frac{dT(x)}{dx} \right] = 0
\]  

(46)

Equation (1) is said to be a macrophysical description of the rod or a “macroequation”. Importantly, \( K(x) \) is “a slowly changing function of position” (Rueger 2006: 338). \( dT(x)/dx \) is the rate of change of temperature at position \( x \). This, multiplied by the rod's conductivity, i.e. \( K(x)(dT(x))/dx \), is the heat flux, or the rate of heat transfer, at position \( x \). The equation is therefore a measure of the rate of change of heat transfer over position, and implies that the heat flux at any point in the rod is a constant function of position.

The “microequation” that is said to explain the macroequation describes a system consisting of individual atoms separated by empty space—a periodic lattice with a period of length \( P = \varepsilon L \), where \( \varepsilon < 1 \):

\[
\frac{d}{dx} \left[ k(x) \frac{dT(x)}{dx} \right] = 0
\]  

(47)
Here, $k(x)$ is the microscopic conductivity. Importantly, $k(x)$ is “a rapidly oscillating function of position (high around the location of the “atoms”, low in the interatomic spaces)” (Rueger 2006: 338).

We want to explain (1) in terms of (2) by deriving (1) from (2). The obvious way of doing so would be to seek a perturbation expansion of the solutions of (2) in terms of a small parameter $\epsilon = P/L$ and expect to retain, in the limit $\epsilon \to 0$, the solution of (1). However, this fails because letting $\epsilon$ go to zero will result in more and more rapid oscillations of the coefficient $k(x)$, and we will not obtain the sort of uniformly convergent series we require. However, one can still construct a uniformly valid approximation in the form of an asymptotic power series by introducing two length scales into the microdescription, the microscopic scale $x$ and a macroscopic scale $X = \epsilon x$. In doing so, we filter out information that previously disallowed us to reach the desired result. However, this requires that we invoke macro information about the macroscopic spatial scale into the microequation in order for the perturbation expansion to approximately recover the solution of the macroequation. Apparently, the micro level cannot by itself explain the macro level, as RET requires.

A number of controversial implications have been said to follow from this example. Rueger (2001, 2006) argues that explanations like this cannot be considered ‘reductive explanations’ in the philosopher’s sense in which “to speak of a theory being reduced is to privilege the descriptions at the reducing level” (2001: 504). This is because the explanation of the features of the macrophysical level require appeal to features of the macrophysical level, and cannot be given solely in terms of features of the microphysical level. Kistler (2006) agrees, but clarifies the specific sense of ‘reductive explanation’ that the example challenges. In particular, Kistler takes the example to undermine views that require inter-level reductions to exhibit derivability from the microphysical level alone, such as RET.

Rueger (2004) uses the example to motivate the idea that instances of macro-structural properties (e.g. macro conductivity) are parts of instances of micro-structural properties (e.g. micro conductivity), that cannot be reduced to those micro-structural properties. He then uses this to solve the causal exclusion problem. This problem (roughly) states that the cause of any event can be fully accounted for in microphysical terms, and that because of this, macrophysical properties play no causal role in the world. Rueger’s solution denies the reductionism assumption: the macro temperature of the rod is caused by the macro conductivity, and these properties cannot be fully accounted for (reduced) in microphysical terms.

Finally, Rueger and McGivern (2010) use the example to defend a new conception of levels of reality. They deny that levels of reality should be individuated in terms of entities, where the lowest level is inhabited by the smallest or simplest entities, and the higher levels inhabited by the larger, complex composites. They argue that levels are best individuated in terms of behaviours of entities rather than
entities themselves. For example, the same entity (e.g. the rod) can exhibit macro and micro “behaviour”. Using the example, they develop the idea that the macro behaviour is a part of the micro behaviour, not vice versa, and they develop an interesting new conception of levels of reality based on this.

If RET is true, then the explanation is either incomplete, or defective. Rather than arguing that it is incomplete, I instead argue that it is defective, at least as Rueger and Kistler have stated the explanation. The problem is this: equations (1) and (2) cannot both be true because they contradict one another. The important difference between them are the terms $K(x)$ and $k(x)$, where the former represents the conductivity of the rod and is a slowly changing function of position, while the latter represents the conductivity of the rod and is a rapidly oscillating function of position. But how can the conductivity of the rod be both a rapidly oscillating function of position and a slowly changing function of position? It simply cannot. So at least one of the equations is false, most obviously, the macro equation.

The reduction is described as “the reduction of the macroscopic thermal conductivity to the microscopic thermal conductivity”, and so perhaps the prefixes “micro” and “macro” can be defined so as to avoid the contradiction. But this does not seem promising. The micro conductivity is not a property of parts of the entity that exhibits the macro conductivity. On the contrary, both are properties of the exact same entity—the rod. But the rod itself just has a particular conductivity; it does not have both “macro” conductivity and a “micro” conductivity.

Rueger and McGivern (2010) anticipate a similar objection: “she [the reductionist] could simply insist that the true description of the macro behaviour is that which comes from an exact, not approximate, solution to the micro theory” (2010: 388). They respond as follows: “Since we don’t have [an exact solution to the micro theory], we haven’t been shown that we can extract a more accurate description of the macro behaviour from it: we’ve simply assumed that we must be able to. Showing this would involve investigating a limit relation of some sort, and we’ve just seen that the results of such operations can’t be taken for granted.” (2010: 388). But this fails to address the key issue. For the reductionist should deny that one needs to extract a more “accurate” description of the macro behaviour from some other description. No such description needs to be extracted because there is no “macro” conductivity to be described.

While equation (1) is false (assuming the truth of (2)), and therefore not a datum to be explained, there are nonetheless truths in the vicinity, worthy of explanation. One such truth is: From the perspective of scale S, equation (1) is empirically adequate. There are a number of ways one might make the notions of “the perspective scale S” and “empirically adequate” precise. One way of spelling them out would appeal to scales in which the results of certain (precise) measurements can be ignored.
Such scales act as information filters. If our scales are spatial scales then at the lowest scale, all microphysical information is accounted for including interactions occurring in the smallest possible regions. At higher scales, information discernible only in small spatial regions is filtered out.

With the explanandum more accurately formulated, we can now ask: does the explanation (and explanations of its type) support or undermine RET? RET advocates can account for these types of perturbation explanations insofar as the explanandum of such explanations mention the scales in which the macro equations are empirically adequate. In particular, from a purely microphysical description, one can derive the existence of various possible scales such as the macroscopic scale $X = \varepsilon x$. After all, if such scales are information filters, then one can in principle take a detailed microphysical description and filter out whatever information one likes. One can define a number of higher level scales and use them in perturbation explanations to derive solutions to the equation for the rod’s conductivity relevant to the various scales. In this way, one derives, from purely microphysical premises, the conclusion that from the perspective of scale $S$, equation (1) is empirically adequate.

### 3.3 Inhomogeneous Reductions

The purpose of this section is to respond to the objection that inhomogeneous reductions do not (even indirectly) exhibit a priori entailment. An inhomogeneous reduction has vocabulary in the explanandum that is not contained in the explanans. Inhomogeneous reductions contrast with homogeneous reductions. An example of a homogeneous reduction is the reduction of mass additivity, which contains the crucial term ‘mass’ in both explanans and explanandum. Many authors think inhomogeneous reductions cannot exhibit a priori entailment just because the explanandum introduces new vocabulary. An example is Diaz-Leon (2010), who tries to illustrate this with the reduction of condensation, to molecular chemistry—which does not itself mention ‘condensation’. In what follows I respond to this objection.

As she builds up to her discussion of solidity, Diaz-Leon (2011: 107) considers a conditional that is similar to the (Dots-to-shapes) conditional from section 2.5. and uses it as part of an argument against AET. She considers the following conditional:

$$(AC_{\text{square}}) \text{ If } x_1, x_2 \ldots x_n \text{ instantiate properties } F_1, F_2 \ldots F_n, \text{ then } r \text{ is square.}$$

Here, $x_1, x_2 \ldots x_n$ are microphysical objects while $F_1, F_2 \ldots F_n$ are microphysical properties. Diaz-Leon provides no further information as to what these variables stand for, other than that the antecedent
makes no explicit mention of squares. The reason the conditional is named \( (AC_{\text{square}}) \) is because the conditional is meant to be an *application conditional* for the concept of a square. This notion comes from Chalmers and Jackson (2001)—Diaz-Leon’s primary target. By definition, an application conditional is a material conditional that satisfies (i) and (ii) while an *a priori* application conditional also satisfies (iii):

(i) **ANTECEDENT:** The antecedent contains a description of some environment that does not contain concept C. For example, if the environment description is a dot-distribution description, we may not want the description to contain the concept SQUARE. To do that the description must not involve the word ‘square’ or synonyms.

(ii) **CONSEQUENT:** The consequent is a proposition containing C, which identifies C’s extension with something described by the environment description. For example, if \( C = \text{SQUARE} \), then our consequent will say ‘x is a square’, where ‘x’ is some part of the antecedent description such as ‘the dots with such and such co-ordinates’. This *extension finding proposition* may also just say ‘there is a square’.

(iii) **EPISTEMIC STATUS:** The antecedent must be rich enough so that the conditional can be believed with a priori justification. In other words, all empirical evidence relevant to the evaluation of the conditional must be loaded into the antecedent, so that the conditional can be justifiably believed a priori.

The conditionals discussed so far such as (Dots-to-shapes) and (Mass additivity), as well as the relevant expansion conditionals are all application conditionals, which I argue are *a priori* application conditionals.

Chalmers and Jackson argue that a wide range of non-trivial application conditionals are *a priori* because of a plausible constraint on concept possession. The claim is that possessing a concept partly involves possessing certain *conditional abilities*. Conditional abilities consist in being able to correctly apply a concept to a new environment, given sufficient information about that environment. They are called *conditional* abilities because having them puts one in a position to evaluate conditionals that have sufficiently detailed environment descriptions in their antecedents and relevant extension finding sentences in their consequents. Our conditional abilities explain why there are a range of non-trivial application conditionals that we know a priori. This can be illustrated with the simple example of the concept SQUARE. Possessing this concept partly involves being able to correctly apply it to certain environments (e.g. those containing lines arranged in certain ways). It is plausible that concept possession comes with conditional abilities: to determine whether subjects possess concepts we determine whether they have the relevant conditional abilities. For example, to determine whether a child has grasped the concept SQUARE, we present them with lines arranged in certain ways and ask them to say what shape is made.
Against the apriority of (AC\text{square}), Diaz-Leon assumes that if one understands the word ‘C’, then one has the concept C, and she defends the claim that there are subjects who understand ‘square’ and therefore possess SQUARE, but who are not in a position to know (AC\text{square}). She considers a hypothetical subject, Tina, “who is not in a position to know the application conditional (AC\text{square})” (2011: 110). Diaz-Leon then asks whether Tina could possess SQUARE despite the apparent lack of conditional ability. Diaz-Leon asserts that “It seems perfectly compatible with lacking such a conditional ability that she [Tina] could in effect use the word ‘square’ very much as other subjects who have the conditional ability would” (p110). Diaz-Leon then summarises her objection as follows:

“We can put the problem this way: if we assumed that in order to possess macroscopic concepts such as SQUARE we have to associate them with reductive application conditionals such as (AC\text{square}), then it would follow that we lack many concepts that we thought we did possess. [...] However, if in order to possess them we had to be able to infer truths involving them from lower-level descriptions, it would turn out that we do not really possess such ordinary concepts, since we do not seem to have such sophisticated inferential capacities”.

(Diaz-Leon 2011: 110)

In response, there are many reasons why a (seemingly) competent speaker such as Tina might not “be in a position to” know (AC\text{square}), consistent with the conditional abilities model of concepts. For example, on one reading of “being in a position to know”, Tina is not in a position to know (AC\text{square}) because she does not possess the concepts in the antecedent. But this is irrelevant to the conditional abilities model. The conditional abilities model is only committed to the claim that if Tina understood and contemplated the antecedent then possession of SQUARE and the ability to reason with it is the only further ingredient required for Tina to know (AC\text{square}).

For Diaz-Leon’s objection to work, the antecedent of (AC\text{square}) needs to be properly specified. Then we need a (plausible hypothetical) subject who fully understands that antecedent, possesses SQUARE (and can reason with it), but is not in a position to know (AC\text{square}). But we have been given no reason to think that such subjects exist. An example similar to (AC\text{square}) is (Dots-to-shapes), which is a priori, and seems a priori partly in virtue of the conditional abilities we associate with SQUARE. If one was presented with the dot distribution, fully understood it, and was given time to categorise the shapes that the dots compose, yet failed to identify the square, then this would be good reason to deny that one possesses SQUARE. And to convert (Dots-to-shapes) into an example that better satisfies the (AC\text{square}) schema, we just need to convert the two dimensional space into a three or four dimensional space, and our two dimensional dots into three or four dimensional mass densities.

\footnote{60 A thesis she has borrowed from Williamson (2003, p290).}
This is one way out of Diaz-Leon’s argument. Another way to make sense of how Tina could be competent with ‘square’ yet not be in a position to know (AC\text{square}), is if Tina is a deferential user of the term ‘square’. That is, Tina associates few conditional abilities with ‘square’ except that she is able to mean by ‘square’ what more competent users of the term mean by ‘square’. That is, although she cannot pick out squares very well, she can nonetheless transmit information about squares to competent users of ‘square’. And if someone challenges her on the information she is trying to pass on, Tina is able to defer the challenger onto the right sources. For Tina, the only a priori application conditionals she associates with ‘square’ are ones whose antecedents describe who the expert users of ‘square’ are and how they apply ‘square’.

Diaz-Leon herself raises this as a possible response to her argument (2011: 111-112), before briefly raising an objection to it, concerning a situation in which there are no expert users of a term. Her concern is that “if there are no experts on a certain matter, it follows that no-one fully possesses the corresponding concepts, and furthermore, that no-one can even have a partial grasp of a concept, since there are no experts to rely on” (p112). But this confuses partially possessing a concept with possessing a deferential concept (note that one can have partial or full possession of both deferential concepts and non-deferential concepts). A deferential concept is a concept whose associated a priori application conditionals all have reference to expert uses of the relevant term in the antecedent. Non-deferential concepts are associated with many a priori application conditionals that don’t reference experts in the antecedents. A given non-deferential concept is more fully grasped by someone the better they are at finding its extension given sufficient information.

In previous sections we understood the concept MASS, and other concepts, in a non-deferential manner. Roughly, possessing MASS requires that one has the conditional abilities to identify mass as a certain property of things that disposes those things to resist changes in motion (e.g. acceleration) in response to certain applied forces (e.g. pushing forces). This is only a rough analysis. That’s because there is no particular reason to think that precise analyses are possible, nor is there reason to think that AET requires them. According to the conditional abilities model, concept possession requires that one be in a position to know certain application conditionals. An analysis of a concept involves taking many of these application conditionals, determining what it is in their consequents that are important to determining the extension of the concept, and generalising over those properties. For example ‘a square is the type of thing that is composed of four lines arranged in such and such a way’. Whether such generalisations yield finitely expressible necessary and sufficient conditions for the application of the concept is not important for evaluating AET. What’s important is whether the application conditionals used to expand a more complex conditional are a priori. Having a rough analysis of the concept in the consequent, given in terms of concepts in the antecedent is just one method for arguing that the application conditional is a priori, and corresponds to Analysis Based Justification from section 1.1.
Diaz-Leon (2010) argues that while conditional abilities model of concept possession may guarantee that homogeneous reductions exhibit a priori entailment, it does not guarantee that inhomogeneous reductions exhibit a priori entailment:

“This might be right for some cases: maybe if we know, for instance, the individual masses of microphysical entities $x_1, x_2 \ldots x_n$, which compose the macrophysical entity $r$, then we can infer the mass of $r$. This seems plausible because we are using the same predicate both at the microphysical and macrophysical level, namely, ‘mass’. But what happens when we introduce new predicates at higher-order levels? A complete description of the macrophysical level seems to require new predicates that did not appear at the microphysical level. Are we really able to apply these novel macrophysical concepts, given a microphysical description, without appealing to further empirical knowledge?”

(Diaz-Leon 2010: 106-7)

Diaz-Leon (following Levine (2001)) refers to the view that concept possession comes with conditional abilities as *Ascriptivism*. She distinguishes two different kinds of ascriptivism, non-reductive ascriptivism and reductive ascriptivism (2010: 108-109). Diaz-Leon’s definition of *non-reductive ascriptivism* is:

“For any concept $C$ at level $n$, there is an application conditional like this: (AC): ‘If $x$ is $F$, then $x$ falls under $C$’, where feature $F$ is described using only concepts at level $n$”

(Diaz-Leon 2010: 109)

Diaz-Leon’s definition of *reductive ascriptivism* is:

“For any concepts $C$ at level $n$, there is an application conditional like this: (AC): ‘If $x$ is $F$, then $x$ falls under $C$’, where feature $F$ is described using concepts from a lower level $m$”

(Diaz-Leon 2010: 108)

Diaz-Leon does not clarify what ‘level’ means, except to say that microphysical concepts (‘mass’, ‘charge’) are concepts at the microphysical level, macrophysical concepts (‘square’, ‘condensation’) are concepts at the macrophysical level, and macroscopic concepts (‘water’, ‘economy’) are concepts at the macroscopic level.

With this distinction in hand, Diaz-Leon rightly notes that:

“We can accept the basic ideas of ascriptivism, and still deny the additional assumptions of *reductive ascriptivism*. For example [...] an advocate of non-reductive ascriptivism does not have to accept that microphysical truths a priori entail ordinary macroscopic truths”

(Diaz-Leon 2012: 108-109)
If we accept non-reductive ascriptivism but reject reductive ascriptivism, then we can deny that sufficient microphysical information a priori entails higher level truths containing new vocabulary. For if all a priori application conditionals associated with higher level concepts have other higher level concepts in their antecedents, then there are no a priori application conditionals that one could appeal to, to justify an a priori inference from microphysical truths alone to a higher level truth. Thus, if reductive ascriptivism is false, then so are RET and AET.

We have already seen that reductive ascriptivism is unproblematic if we take our higher level concept to be ‘square’, and our lower level descriptions to describe dot-distributions (section 1.5). But perhaps when we move to more realistic cases, involving the microphysical and macrophysical level, reductive ascriptivism is more problematic. Diaz-Leon believes that the concept of condensation causes problems for reductive ascriptivism. She notes that the relevant application conditionals associated with the concept would take the following form:

\[(\text{AC}_{\text{condensation}}) \text{ If } x_1, x_2 \ldots x_n \text{ instantiate properties } F_1, F_2 \ldots F_n, \text{ then } r \text{ condenses.}\]

Diaz-leon’s goal is to argue that once we are clear on what the possible values for the variables \(x_1, x_2 \ldots x_n\) and \(F_1, F_2 \ldots F_n\) are, given reductive ascriptivism, we will see that \((\text{AC}_{\text{condensation}})\) cannot possibly be a priori. Her argument involves a subject who learns the reductive explanation of condensation:

Let’s now consider a subject, Bette, who is at the beginning of a degree in chemistry. During her first weeks at college, she does not know yet what microphysical properties determine the condensation of vapour. That is, she does not know \((\text{AC}_{\text{condensation}})\). Therefore, according to the view we are considering [reductive ascriptivism], Bette does not fully possess the concept CONDENSATION, nor does she fully grasp the meaning of the expression ‘condensation’: she has to learn more facts in order to fully grasp the concept and understand the term.

(Diaz-Leon 2010: 112-3 [my italics])

Notice the italicised sentence: Diaz-Leon is assuming that if reductive ascriptivism is true, then the values of the variables \(F_1, F_2 \ldots F_n\) must be the microphysical properties that actually determine condensation. But this is a problematic assumption: the conditional abilities we associate with ‘condensation’ are supposed to enable us to infer the existence (or non-existence) of condensation given arbitrarily many microphysical descriptions with distinct condensation realisers. Diaz-Leon continues:

“This is the aspect of this position that I find problematic: it is held that certain knowledge is a priori for a subject, but at the same time we have seen that this subject has to acquire that knowledge in a way that resembles paradigmatic a posteriori knowledge, such as going to chemistry lectures. [...] But how could such knowledge suddenly become a priori knowledge?
Going to chemistry lectures seems to be a paradigm of coming to know something a posteriori. [...] The problem here is that there seems to be no way that we could come to know the information that is relevant in order to know the corresponding application conditionals in an a priori manner. For how could we come to know that such and such microphysical properties determine a body’s condensation, merely a priori?"

(Diaz-Leon 2010: 112-113 [my italics])

If Diaz-Leon is right about what reductive ascriptivism entails: that the actual properties that determine condensation must be known in order to possess the concept ‘condensation’, then reductive ascriptivism is implausible. However, this cannot be right. In fact, there appears to be a conflation between the following two claims:

**(Condensation) Claim one:** Reductive ascriptivism requires that if Bette possesses CONDENSATION then Bette knows that such and such microphysical properties determine a body’s condensation.

**(Condensation) Claim two:** Reductive ascriptivism requires that if Bette possesses CONDENSATION then Bette is in a position to know that if such and such microphysical properties obtain then they determine a body’s condensation.

There are two important differences between claim one and claim two. Firstly, claim one concerns knowing whereas claim two concerns being in a position to know. Secondly, claim one concerns knowing that certain properties determine condensation whereas claim two concerns knowing the conditional that if certain properties obtain then they determine condensation. Diaz-Leon’s argument assumes claim one. However, claim one is a misinterpretation. Reductive ascriptivism only requires claim two. To see this, consider the analogue in the dots-to-shapes example (section 1.5):

**(Square) Claim one:** Reductive ascriptivism requires that if Bette possesses SQUARE then Bette knows that such and such a dot-distribution determines a square.

**(Square) Claim two:** Reductive ascriptivism requires that if Bette possesses SQUARE then Bette is in a position to know that if such and such a dot distribution obtains then there is a square.

We can similarly illustrate the point for mass additivity:

**(Mass) Claim one:** Reductive ascriptivism requires that if Bette possesses MASS then Bette knows that such and such a fundamental law determines mass additivity.

**(Mass) Claim two:** Reductive ascriptivism requires that if Bette possesses MASS then Bette is in a position to know that if such and such a fundamental law obtains then mass is additive.
Clearly, claim two better fits my overall defence of AET/RET. Thus, possession of ‘mass’ requires that one is in a position to know that if a number of particles is governed by Newton’s laws then mass is additive. One need not know this conditional to possess ‘mass’. Rather, one only needs to know that the mass of an object is its most natural disposition to resist changes in motion given applied forces. As argued, knowing this puts one in a position to know the conditional. For if one is presented with the relevant microphysical information, then possession of the relevant concepts enables one to determine how ‘mass’ applies.

To possess ‘condensation’ one simply needs to know that an object’s condensation is its most natural property responsible for its change from a vapour state to a liquid state. So if one accepts non-reductive ascriptivism, as Diaz-Leon does for the sake of argument, then this should not be problematic. Diaz-Leon, however, anticipates this response:

“[P]erhaps the empirical facts about condensation that we are referring to are not relevant here: what is relevant is the causal functional role that we associate with ‘condensation’, and not the microscopic properties that happen to realize that role in the actual world. [In response, the] thesis of a priori entailment requires that we are able to apply some macrophysical concepts to a microphysical description, and in order to do that it is not enough that we know the causal-functional roles associated with those macrophysical concepts: we also have to be able to apply these roles to a microphysical description, and it seems that we need to be able to apply some macrophysical concepts to a microphysical description, in order to do that.”

(Diaz-Leon 2010: 113 n11)

Diaz-Leon is right: to apply the causal functional roles, we need to first know that there are some macrophysical objects playing those roles. But how do we infer their existence from a purely microphysical description? To answer this question in the cases of (Dots-to-shapes) and (Mass additivity), we just need to look to the expansion conditionals. For it is the intermediary expansion conditionals that set us up for the application of the relevant causal functional roles. What does most of the work is the [Composition] expansion conditional. In (Dots-to-shapes) we simply needed to group the dots into composites, in order to find lines arranged in certain ways. That was enough to apply the “square-role”—because the four lines were playing it. Similarly in (Mass additivity) a composition premise allowed us to find the position and acceleration of composites. Furthermore, because the microphysical description also described total forces, we were able to infer composite forces. This gave us enough information to apply the causal-functional roles that we associate with our two notions of mass. Now, insofar as parts and wholes represent different levels, and insofar as total forces and composite forces represent different levels, then we have located where the crucial reductive-ascriptivism assumptions lie. But they seem unproblematic: we simply do not learn these
micro-macro inferences from going to science classes. The ability to infer composites from parts simply comes from possessing ‘composite’ (section 2.8) and inferring composite forces from total forces similarly flows from grasping the concepts of ‘total force’ and ‘composite’.

Let’s extend this reasoning to Diaz-Leon’s case study. If ‘condensation’ is best analysed in terms of the most natural property responsible for change from a vapour state to a liquid state, then an a priori entailment expansion will require expansion conditionals that infer, from the relevant microphysical base, vapour and liquid. Expansion conditionals taking us from microphysics to the various states of matter such as vapour and liquid, will themselves need to be expanded to be properly analysed. For this reason I will put forward an a priori entailment expansion for the simplest state of matter: solidity. I will then argue that because such an expansion is possible, so are expansions for vapour and liquid. From there, we can argue that expansions for changes between such states of matter, and hence, condensation, is also possible.

Let’s begin with Jackson’s instructive (1998) remark about finding solidity:

“How Consider the story science tells about tables, chairs, pens and the like being aggregations of molecules held in a lattice-like array by various intermolecular forces. Nowhere in this story is there any mention of solidity. Should we then infer that anyone who thinks that the story science tells us about these dry goods is, in some sense, a complete story, is committed to nothing being solid? Obviously not. The story in favoured terms will, we may suppose, tell us that these lattice-like arrays of molecules exclude each other, the intermolecular forces being such as to prevent the lattices encroaching on each others spaces. And this is what it takes, according to our concept, to be solid. […] Hence, solidity gets a location or place in the molecular story about our world by being entailed by that story”

(Jackson 1998: 3-4)

This is the beginnings of an a priori entailment expansion. Let us finish the job by expanding the following conditional:

(Solidity): If microphysical description & [T] & molecular description then composite one (C1) is more solid than composite two (C2).

As always, what we get out of our antecedent is a function of what we put into it, so let us begin by being clear about exactly what it includes. Our microphysical description will be similar to those previously described, except that it will contain many more particles. Furthermore, while previous microphysical descriptions eliminated time—considering worlds consisting in single time-slices, the present world can consist in a short time period, let’s say one minute. The microphysical description will also describe many more laws of nature, including laws that describe forces which hold atoms
together. We shall add to this microphysical description a “that’s all that is fundamental” stop clause as usual (so it is a complete description of what’s fundamental). We shall also add a comprehensive molecular description.

Our molecular description will do the primary work in a priori entailing the consequent of (Solidity). It describes the properties of 2000 molecules \( \{M_1...M_{2000}\} \) as they undergo a brief period of existence. It will describe the intrinsic properties \( \{I_1 ... I_{2000}\} \) of each molecule at every point in time, properties such as mass and being composed of specific covalently bonded atoms. It will describe the laws \( L \) that relate the molecules in various ways, such as through electrostatic attractions and van der Waals forces. It will also describe in full detail the trajectories \( \{t_1 ... t_{2000}\} \) of all the molecules. The antecedent is being described in such a way so that (i) it follows apriori that two composites undergo a collision where one of the composites shatters and (ii) this post-collision shattering is a result of the initial positions of the composites and laws governing their trajectories. The thought is that this will be enough to allow us to infer a priori that one of the composites is more solid than the other.

With this in mind let’s expand (Solidity):

(1) [Microphysical/macrophysical description]:

[Microphysical description] & [T] & There are two thousand molecules \( \{M_1...M_{2000}\} \) that have intrinsic properties \( \{I_1 ... I_{2000}\} \) and are governed by laws \( L \). \( M_1 \) to \( M_{1000} \) have trajectories \( \{t_1 ... t_{1000}\} \) such that they are initially located in the North West and are travelling together towards the South East. Molecules \( M_{1001} \) to \( M_{2000} \) have trajectories \( \{t_{1001} ... t_{2000}\} \) such that they are initially located to the South East and are travelling together towards the North West.

(2) [Composition] If (1) then there are two composites —\( C_1 \) composed of \( \{M_1...M_{1000}\} \) and \( C_2 \) composed of \( \{M_{1001} ... M_{2000}\} \)—both composites are located where their parts are located at any given time.

(3) [Encroachment] If (1) and (2) then \( C_1 \) and \( C_2 \) collide, \( C_2 \) deforms (behaves brittlely) while \( C_1 \) resists the encroachment of \( C_2 \) on its space (behaves solidly).

(4) [Robustness] If (1)-(3) then \( C_1 \)’s solid behaviour and \( C_2 \)’s brittle behaviour are robust over nomological possibilities in which \( C_1 \) and \( C_2 \) collide.

(5) [Solid] If (1)-(4) then \( C_1 \) is more solid than \( C_2 \).

(6) [Macroscopic description]: \( C_1 \) is more solid than \( C_2 \).
I will now defend the apriority of premises (2) to (4) and hence the apriority of (Solidity).

Premise two concerns the issue of what it takes for some objects to compose a composite. I have already discussed this issue in detail in section 2.8, and won’t repeat that discussion, except to make a few observations. Firstly, we have a priori inferred just two composites, but there are clearly many more than just two. Furthermore, the two we have chosen—$C_1$ and $C_2$—do not change their parts over time. We could just as well run our argument with composites that change their parts over time but it will involve further complexities that we need not get into. Thus, by grouping together certain particles (at each time) in thought and then applying names to them, we a priori infer the existence of $C_1$ and $C_2$ and we a priori infer where $C_1$ and $C_2$ are located at each moment in time, and therefore, their complete trajectories.

The consequent of premise three [Encroachment] introduces several new macrophysical terms including ‘encroachment resistance’ and ‘deformation’. But hopefully it is clear that the applicability of such notions should be transparent to anyone who understands these notions and understands the antecedent. After comprehending premise (1), a suitably equipped reasoner should be able to simulate the minute long event in imagination, thereby simulating two collections of molecules (each containing one thousand molecules) colliding. The simulation will make clear that the components of $C_1$ collectively hold their ground in response to the collision while the components of $C_2$ ping off in all different directions, causing a drastic shape change in $C_2$, or a “deformation”.

From the fact that a collision gave rise to solid behaviour in $C_1$ and brittle behaviour in $C_2$, it does not follow that $C_1$ is more solid that $C_2$. For this may have just been a coincidence in which the more brittle $C_1$ just so happened to hit a weak point in the more solid $C_2$ triggering a chain reaction of breakages in other weak points of $C_2$, coincidently causing it to break up. So one needs to rule this hypothesis out. This is where premise (4) [Robustness] comes in. Here we simulate nomological possibilities, tweaking the initial conditions and then applying the fundamental laws to simulate the various post-collision trajectories of $C_1$ and $C_2$. The idea is that (1) is spelled out so as to guarantee that the post-collision result similar across a variety of variations in initial conditions. This rules out the coincidence objection and allows us to postulate dispositions to exhibit solid behaviour and dispositions to exhibit brittle behaviour, in response to such collisions. This brings us to premise (5), in which we infer that $C_1$ is more solid than $C_2$.

The idea that being disposed to resist encroachment is sufficient for solidity, is based on the simple idea that if you tap your knuckles on an object, and it resists the encroachment of your knuckles upon its space, it is solid. Of course, being disposed to resist encroachment is not a strict analysis of solidity given the availability of counterexamples. For example, metal is not disposed to resist the encroachment of acid upon its space: given enough time, acid will deform the metal. But intuitively, the acid is not more solid that the metal. This is all derivable from the microphysics given that forces
quite different from collision forces are involved. So the antecedent description will rule out such
defeateds. And so we get to a conclusion about relative solidity and nowhere did we rely upon obvious
or paradigmatic a posteriori conditionals.

One might worry that while we have inferred *something like* solidity, we haven’t inferred *solidity.*
Perhaps objects are either solid or they aren’t (setting aside borderline cases), whereas we have only a
priori inferred *relative* solidity. That is we have only inferred that one object is more solid than
another without inferring whether or not the objects are solid in themselves. I think the correct thing
to say here is that as with most words, the word ‘solidity’ is highly polysemous in the sense that it is
ambiguous between a number of related concepts. Thus, I think what has been achieved so far is that
a notion of solidity has been shown to give some confirmation to reductive ascriptivism and
AET/RET. But I shall briefly consider related notions of solidity to see if they might give rise to any
problems.

We can roughly individuate different notions of solidity by giving approximate definitions of them.
Thus, Jackson’s notion of (relative) solidity can be approximately defined in terms of one object being
more solid than another if it is disposed to resist the encroachment on its space of the other object in a
collision. An absolute notion of solidity might be approximately defined in terms of *being everywhere
dense* such that if an object has even the slightest hole in it then it is not completely solid. I’m not sure
if anyone has ever in fact possessed this concept, but one explanation for why some people believe
nothing is solid because of the atomic discovery that tables (etc.) are mostly empty space, could be
that they associate this concept with their term ‘solidity’. The more important point to note is that it is
extremely plausible that truths involving this macroscopic concept follow a priori from microphysics:
just group matter up into composites and ask whether or not the matter in those composites are
separated by space. If this is never the case then one will be in a position to infer a priori that no
things are solid (in this sense of solidity).

One way of using solidity to object to reductive ascriptivism would be to argue that Jackson’s
approximate analysis in terms of encroachment resistance is on the right track, but that “the real”
notion of solidity is tied to being disposed to resist the encroachment of an object on its space given a
certain *prototypical* kind of applied force. In particular, the concept of solidity is a prototype concept,
where the prototypical case of a solid object is an object that is disposed to resist the encroachment of
an *ordinary human’s knuckle* upon its space when such a human taps her knuckles on it. Everything
else is judged to be solid in terms of some similarity to the prototypical solid object. Such concepts do
raise problems for AET. For how could we infer whether or not something is solid in this sense
without first inferring humans, their knuckles, and the things that resist knuckle-taps? If we can first
infer humans and their knuckles from microphysics this is fine, but a problem will arise if in order to
infer the existence of humans, we first need to infer what things are solid and what things are not solid.

The correct response is to again appeal to the polysemy of the word ‘solid’. For while it’s true that we first need to infer many human-truths before being able to infer these solidity-truths, there are various non-anthropomorphic notions of solidity that can help us to infer humans-truths. In fact, this is precisely the reason why AET/RET seems so plausible to me: there are just so many concepts in conceptual space whose extensions can be inferred from microphysics. Once their extensions are inferred we are given an enormous base of information from which we can a priori infer the extensions of our more familiar folk concepts. Think about all the concepts that can possibly be formulated through rigorous empirical enquiry, or introduced by the roles they play in theories. For any concept tied to humans as this disambiguation of ‘solidity’ is, there will be strongly related non-anthropomorphic concepts that will help us to a priori infer humans (or whatever high-level properties our ordinary notions are constitutively tied to).  

Allow me to illustrate this point with the notion of solidity invoked in modern rheology (‘the study of flowing things’). A much more precise and objective notion of solidity is approximately defined in terms of how an object is disposed to behave in response to a stress, applied by a force with a given magnitude, over a given duration. The size of the force generally determines the size of the response: the larger the applied force, the larger the deformation.

Most objects exhibit solid-like behaviour (e.g. by deforming elastically or in a brittle manner) given forces applied over short time-scales, but exhibit fluid-like behaviour given forces applied over long time-scales. We can define these notions more carefully as follows:

Solid-like behaviour: total deformation is a function of the magnitude of the applied force.

Fluid-like behaviour: deformation rate is a function of the magnitude of the applied force.

A solid will only deform a fixed amount under an applied stress, while a fluid will continue to deform until the stress is removed. Furthermore, a solid will usually rebound if it hasn’t undergone brittle deformation once the stress is removed, while the deformation to the fluid is permanent. We can now approximately define solid objects as follows:

---

61 Relatedly, Levine (2010) argues that all concepts are anthropomorphic such that they are in some way dependent on causal or counterfactual relations to phenomenal consciousness, and he uses this to argue that one cannot a priori any truths from microphysics until one has inferred phenomenal truths. But there is little hope of inferring phenomenal truths from microphysical truths and so AET fails. I evaluate this objection in section 3.4.

62 Thanks to Jess Robertson for suggesting this example.
Solid object: an object is solid (relative to the force and timescale) if the force applied is of the right size and timescale for the resulting deformation to be (a) small enough that the object does not fail in a brittle manner, and (b) recoverable when we remove the stress.

To illustrate this concept in action let’s consider some cases. Take a Maxwell body, such as glass. Glass flows like a fluid for small forces over long timescales (gravitational effect on window panes), is solid for small forces over short timescales (knuckle tap), and is brittle for large forces over short timescales (throwing bottles). Now consider a continental plate, which is solid when we walk on it (small force, small timescale), elastic in response to an Earthquake (large force, medium timescale), brittle in response to an asteroid (large force, small timescale), but fluid on the timescale of the age of Earth. With these examples in mind, we can analysis this concept using a rheological plot:63

What this illustrates is that the rheological concept of solidity, while completely non-anthropomorphic, is still non-absolute. In particular, according to this concept, talk of an object being solid is only meaningful when relativised to the magnitude of an applied stress and a timescale over which it is applied.

Hopefully it is clear that possession of this concept enables one to find its extension given enough microphysical information. In particular, we can look to a microphysical description, group whichever particles into composites we please, and then consider how those particles collectively behave in response to an applied stress over a period of time. If the force applied is of the right size and timescale for the resulting deformation to be (a) small enough that the object does not fail in a brittle manner, and (b) recoverable when we remove the stress, then the object is solid relative to that force and timescale. It is hard to see how any other concept of solidity would be required to a priori infer

63 Image courtesy of Jesse Robertson.
the existence of humans. Hence, this concept of solidity is the key macroscopic concept that would mediate the inference to truths involving anthropomorphic solidity concepts.

Here I have focused on the a priori entailment of solidity. But all the same arguments apply equally to fluidity and also to being gaseous. But if that’s right, and we can a priori infer what objects are solid, brittle, liquid, and gaseous, then we should also be able to a priori infer changes between these properties. That is, if we look to the microphysical description and see that particles compose an object that has evolved quickly from a gaseous state into a liquid state, then we are in a position to deduce the existence of condensation. This is how to a priori deduce the phenomenon that troubled Diaz-Leon. And nowhere did the inference require that we treat any paradigmatic a posteriori truths as a priori.

3.4 Reduction of Consciousness

The most common application of AET/RET is in the philosophy of mind, where the hope is that AET/RET can shed some light on why phenomenal consciousness is so hard to explain. An entity is phenomenally conscious if and only if there is something it is like to be that entity; and a mental state is phenomenally conscious if and only if there is something it is like to be in that state. A phenomenal truth is a truth about a phenomenally conscious entity or state.

It is widely believed that phenomenal truths do not follow a priori from non-phenomenal truths. In this section, I consider an objection, which accepts much of the AET/RET framework, but denies that we can infer the fundamentality of consciousness from the fact that phenomenal truths do not follow a priori from non-phenomenal truths.

Let's begin by looking at how one can appeal to AET/RET to argue for the fundamentality of consciousness.

The AET argument for the fundamentality of consciousness:

(1) **AET**: If there is no a priori entailment from fundamental non-phenomenal truths to phenomenal truths then phenomenal truths are fundamental.
(2) There is no a priori entailment from fundamental non-phenomenal truths to phenomenal truths.
(3) Therefore, phenomenal truths are fundamental.

---

64 This is sometimes called "the hard problem of consciousness" (Chalmers: 1995).
The RET argument for the fundamentality of consciousness:

(1) **RET**: If the explanantia of consciousness reductions do not a priori entail their explananda (phenomenal truths), then consciousness reductions are either incomplete or defective.

(2) The explanantia of consciousness reductions do not a priori entail their explananda.

(3) Consciousness reductions are not incomplete: no amount of explanans enrichment enables an a priori entailment expansion of ‘If [non-phenomenal description] then [there is consciousness].’

(4) Consciousness reductions are defective (from 1, 2, and 3).

(5) If consciousness reductions are defective then phenomenal truths are fundamental.

(6) Phenomenal truths are fundamental (from 4 and 5).

These arguments illustrate how AET/RET can be used to make substantive metaphysical claims. A variety of strategies have arisen as reactions to such arguments. The major reactions illustrated in the following table:

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>A Posteriori Reductionism</em>: Reject AET/RET.</td>
<td><em>A Posteriori Physicalism</em>: Phenomenal truths are grounded in physical truths, but do not follow a priori from physical truths.</td>
</tr>
<tr>
<td><em>A Priori Reductionism</em>: Accept AET/RET.</td>
<td><em>A Priori Physicalism</em>: Phenomenal truths are grounded in physical truths and follow a priori from physical truths.</td>
</tr>
<tr>
<td>Dualism: Explanandum is irreducibly fundamental.</td>
<td><em>Epiphenomenal Dualism</em>: Phenomenal properties are causally inert fundamental properties, caused by fundamental physical properties.</td>
</tr>
<tr>
<td>Eliminativism: Explanandum is an illusion.</td>
<td><em>Eliminativist Physicalism</em>: There are no phenomenal properties—physical truths are fundamental and consciousness is an illusion.</td>
</tr>
</tbody>
</table>

In what follows I consider a novel defence of a posteriori physicalism.\(^{65}\) It appeals to the unique features of our concept of consciousness, to avoid the conclusions of the AET/RET arguments. The objection comes from Joseph Levine (2010), who distinguishes *conceptual fundamentality* from *metaphysical fundamentality*. Levine argues that the AET/RET arguments equivocate on these two notions. He argues that we can explain why consciousness finds the place it does in the AET

---

\(^{65}\) The names 'a priori physicalism' and 'a posteriori physicalism' come from Jackson (2006b). These views are sometimes referred to as 'type-A materialism' and 'type-B materialism', respectively (Chalmers 1996: 165-8).
framework in terms of the concept of consciousness and its relation to all other concepts, rather than in terms of consciousness itself.

While I wish to stay neutral on whether phenomenal truths follow a priori from non-phenomenal truths, it will be worth saying something about why this is widely disbelieved. Take a simplistic object avoiding robot, of the sort that might be built in an undergraduate computer science course. Ask yourself whether you can formulate a clear and distinct idea of it existing without being conscious, that is, without there being something it is like to be the robot. This seems easy. Now enhance the program, as well as the mechanical body that the program controls, so that the robot comes close to being a functional duplicate of a human—having the ability to play chess and pass the Turing test, etc. But do this piecemeal by adding a small bit of functional complexity each day, over several decades. In the process, ask yourself each day, whether you can form a clear and distinct idea of the robot (holding fixed its physical/functional properties) not being conscious. Surely no particular addition to the robot will change things. The robot’s lack of consciousness will remain conceivable, even when we enable it to access its internal states (e.g. its memory) and report them (e.g., on a screen as laptops do). Now replace the robot parts, bit by bit, with living tissue and living organs. Again, each day, it should remain perfectly conceivable that the robot/organism (or whatever it is) not be conscious. Indeed, bring the physical system to the point where it is more or less a physical/functional duplicate of us humans, and describe it on any level you want—quantum, chemical, biological, neurological. The conceivability of its not being conscious should remain. Hence, no a priori application conditional can be formulated in which the antecedent contains purely physical terms, and the consequent only contains a statement like ‘something is phenomenally conscious’.

Let us now look to Levine’s objection to the AET argument. Levine is a physicalist and therefore identifies the fundamental truths with the physical truths. So for Levine, arguments for AET require arguments that some physical truths a priori entail some non-fundamental truths. However, Levine claims that no such evidence has been put forward. In particular, he analyses Chalmers and Jackson (2001), and finds only arguments that non-fundamental truths follow a priori from physical and phenomenal truths. But for Levine, we have no reason to believe in dualism unless we already believe AET, so Chalmers and Jackson's defence of AET is question begging.

One could respond that Chalmers and Jackson's argument provides some evidence for AET whether or not we assume the fundamentality of phenomenal truths. However, Levine has a response:

“But it would be a mistake to interpret their choosing to make their case for the [apriority] of ‘If PQTI then N’ over ‘If PTI then N’ merely on grounds of ease of argument. In fact, from the point of view of their own theory, there is no reason to think that ‘If PTI then N’ would be a priori. What’s supposed to be a priori is the connection between the application of the concept in question and the satisfaction of its reference-fixing conditions. [...] But it’s obvious
that if such a priori reference-fixing conditions exist, they must include facts about how the macro objects and kinds in question affect us phenomenally. If water’s appearance, taste, texture, and the like are not part of our a priori designation of the water role, then what is? But this means ‘If PTI then N’ couldn’t be a priori”.

(Levine 2010: 375)

Before critically evaluating the claim which Levine deems obvious, let’s ask why Levine thinks that Chalmers and Jackson’s own theory commits them to the claim that a priori application conditionals always require phenomenal truths in their antecedents. Chalmers and Jackson claim the following:

“If phenomenal truths are not implied [a priori entailed] by PTI, then it is likely that many other macroscopic truths are not so implied either. For example, knowing whether an object is red arguably requires knowing whether it is the sort of object that causes a certain sort of colour experience, and knowing whether an object is hot arguably requires knowing whether it is the sort of object that causes experiences of heat. If so, then if truths about colour experience and heat experience are not implied by PTI, truths about colour and heat are not implied either”.

(Levine 2001: 318-9 [my italics])

On the face of it, Chalmers and Jackson only commit themselves to the claim that many other macroscopic truths require the addition of Q to PTI in order to be derived a priori—not all macroscopic truths, and certainly not all macrophysical truths. Still, Levine’s idea is presumably that one can keep running the argument that Chalmers and Jackson have started in the above quote. Thus, if colours are not a priori entailed without Q then the same is true of the transparency of matter. But then the existence of water (glass, etc.) may not be a priori entailed either. This is Levine’s objection.

Levine develops his objection further by appeal to a distinction between metaphysical and conceptual fundamentality. He argues that because the application conditions for all concepts depend on the concept of consciousness, the concept of consciousness is conceptually fundamental even though it is not metaphysically fundamental. And he argues that inferring metaphysical fundamentality from conceptual fundamentality simply equates the two distinct concepts.

In this dissertation, I have argued for the apriority of a number of instances of ‘If PT then N’. My method involved breaking such conditionals down into expansion conditionals and arguing for the apriority of the expansion conditionals. Let’s look to see if I was implicitly relying on Q in some way. Then, I will come back to Chalmers and Jackson’s examples of heat and colour.

The initial expansion conditional in all of my a priori entailment expansions concerned composition. Thus, we have some simple objects (described by P) and we a priori infer the claim an example of a
non-fundamental truth N: that there are some composites. Did the argument for this instance of ‘If PT then N’ somehow rely on the apriority of ‘If PQTI then N’ for some Q? It is hard to see how. The nice thing about the composition expansion conditionals is that the epistemic method one employs to infer the consequent from the antecedent is very transparent and easy to describe. One simply groups simples in thought, considers their union, formulates a name, and applies the name to the simples ‘composite sixteen’. The method looks obviously a priori—or at least, it is difficult to see what empirical premises one relies on to justify one’s inference. But Q plays no justificatory role in the inference at all. For surely it matters not whether the inferred composite affects one’s phenomenal consciousness.

In itself this is an insufficient response to Levine because Levine can just say that composition conditionals are a special case that involve, as already noted, a relatively transparent and straightforward a priori inference mechanism. Thus, Levine can simply apply his objection to all other expansion conditionals. The next expansion conditional involved the concept of force. And here, Levine’s objection might have some bite.

Let’s start by getting clear on the role played by the concept of force in both the (Mass additivity) expansions and the (Solidity) expansion. We inferred composite mass from the fact that the composite behaved in certain ways given certain forces (whether gravitational or inertial) and we inferred solidity from the fact that the composite’s were disposed to resist encroachment (or to exhibit solid-like behaviour) given certain forces. But how did we infer the existence of forces? That was easy: forces were already stipulated in the base—some forces are fundamental, and the rest are grounded in (follow a priori from) them. But perhaps this is where Q sneaks its way into the a priori entailment base: perhaps we cannot understand force without consciousness: perhaps 'force' is in some sense a phenomenal notion.

Let’s try to make sense of how this could be. In their book The Concepts of Science, Motz and Weaver (1988) provide an instructive discussion of the concept of force, why it is prior to and partly constitutive of the concept of mass, and where our concept of force first came from:

“[T]he concept of mass as a measurable entity is unfamiliar to us. When we look at an object, we are at once aware of its dimensions, but we can tell nothing about its mass or how its mass is to be measured; nor do we know the nature of mass. [...] [W]e should introduce some quantity other than mass as the third of our basic physical elements. This third quantity should be as perceptible and as easy to estimate by means of our senses as are lengths and intervals of time. Moreover it should lead us to mass as a derived or defined quantity. We do not have far to go to find such a quantity, for it is one that we are constantly involved with, namely force. [W]e are constantly subjected to and must constantly exert forces. Indeed, most of our bodies consist of structures (muscles) whose sole purpose is to exert forces, and the remaining
parts of our bodies are designed to support these muscles or to direct them in their activities. Owing to this constant muscular activity, we have acquired a fairly precise sense of the magnitudes of the forces that are involved in our daily activities. [...] Force may properly be introduced as the third of our undefinable basic quantities” (pp.91-92).

According to Motz and Weaver, we are directly acquainted with three basic indefinable quantities, lengths, intervals of time, and force. From these quantities, other important quantities can be “derived”. For example, mass becomes a property of physical objects crucial to relating applied forces with induced accelerations (acceleration being definable as a time derivative of position change). In that case, mass is indirectly a phenomenal notion if force is a phenomenal notion. So is the concept of force only graspable through grasp of phenomenal effects? And is this also the case for the properties of length and duration?

Even if we answer yes to these questions, we are not then committed to Levine’s premise that all a priori application conditionals for ‘force’ have phenomenal notions in their antecedents. We need to distinguish the properties that fix the reference of a concept from the properties that enabled us to know what properties fix the reference of the concept. It’s true that we feel forces acting on ourselves and we feel ourselves applying forces too. But that doesn’t mean that we cannot know that some object “feels a force” (i.e. has a force on it) without first knowing whether having that force acting on oneself would feel some way. Physicists look to particle collisions and infer inter-particle forces—they do not make such inferences by considering what it would feel like to have such particles collide with oneself (due to their miniscule size one would feel nothing).

Indeed contrary to Chalmers and Jackson’s claim (quoted above) almost no macroscopic truths are such that knowledge of them requires knowing how things are qualia-wise. For all we know, human colour-qualia varies wildly in response to the detection of physical surfaces. Nothing in the modern science of perception conclusively rules this possibility out. But this hardly matters for the communication of colour facts. Take a ripe tomato and a fire truck. We call both of their surfaces ‘red’. But it might be that when I view these surfaces, the colour experience I have is the same as the colour experience you have when you view grass and cucumber, i.e., things whose surfaces we both call ‘green’. So the specific phenomenal responses subjects have to seeing colours are irrelevant to the successful application of colour terms. Rather, whether or not a surface is red depends more on whether or not it is the sort of surface that causes normally functioning English speakers to call it ‘red’. And as for heat, even if an object’s being hot depends on what sort of experiences the object is disposed to cause in humans that touch it (which is debatable), we can simply define up a cousin-concept whose application to objects depends more on whether or not those objects are, for example, disposed to melt and burn things. We can then evaluate AET by deducing truths involving these cousin-concepts, from PT.
A potential line of defence for Levine comes from indexicals. In section 1.1 I noted that indexical truths are both non-fundamental and not a priori derivable from fundamental truths. So indexical truths look like counterexamples to AET. I also noted that throughout chapter two, the worlds I consider when evaluating AET do not satisfy any indexical propositions because they do not contain self-locating agents. Propositions containing concepts whose application depends on indexical facts are therefore either false or indeterminate in such worlds. For example, consider a world containing H₂O manifesting watery properties and XYZ manifesting watery properties, and consider the concept of water which Jackson (1998) analyses as applying to the local watery stuff around here. Since there are no agents in this world for which either XYZ or H₂O could be local to, this concept of water either has no extension in this world or has indeterminate extension. But the polysemy of the term ‘water’ means that there are equally legitimate disambiguations of ‘water’ without indexical elements. There is no reason why propositions containing cousin concepts i.e. non-indexical water concepts—such as the one that applies to the watery stuff simpliciter—cannot be derived from the fundamental description of such a world. For these reasons I set the indexicals problem aside. But the problem needs to be dealt with here. For one might think that indexical truths are another illustration of Levine’s point: indexical truths are conceptually fundamental but clearly not metaphysically fundamental.

Chalmers (2010: 408) argues that indexical truths are unproblematic exceptions: from the fundamental truths (and a “that’s all” clause) one can a priori infer that there are indexicals truths. After all, one can (arguably) a priori infer the existence of self-locating organisms that can and do utter self-locating truths about themselves. For Chalmers, a genuine exception is one where we cannot infer that there are any truths of the relevant type at all. As he puts it, “It is an objective truth that there are inscrutable indexical truths, and this objective truth (like all others) is itself scrutable from fundamental truths. So a basic thesis of scrutability from fundamentals can itself explain the existence of this exception.”

I think this response points in the right direction, but is insufficient. What’s needed is further explanation as to why it should matter to fundamentality that the non-deducibility of a truth is itself deducible. For example, if it turns out that (contrary to common belief) the non-deducibility of phenomenal truths is itself deducible from physical truths, why would that convert dualists into physicalists? Here’s an answer: the phenomenal is not fundamental, provided that you can blame the deducibility failure of specific phenomenal truths, on some feature of the semantics of phenomenal terms (rather than their referents). The key difference is that while being in pain (for example) is a real property of objects; being here, being now, and being me, are not real properties of objects. We pretend that some region has the property of being here for pragmatic (navigational) purposes. But in doing so we know that the region doesn't have some property over and above the non-indexical properties. So even if “Canberra is here” or “I am Kelvin” is not deducible from physics, no property
of the world is missed when we deduce what is deductible from physics. And furthermore, we know
this a priori from our grasp of the concept of here. No such story can be told for phenomenal
properties, which are obviously real properties.

What, then, should we make of Levine’s distinction between metaphysical and conceptual
fundamentality? One can view my a priori entailment expansions as an attempt to show by example
that metaphysical and conceptual fundamentality are closely correlated. But I think it more productive
if we take a sceptical stance toward the distinction. After all, the more closely these notions align the
less clear it seems that anything is achieved by distinguishing them.

Levine speaks as if there are many levels of reality: there is the fundamental level consisting of
physical objects and their physical properties, and then there are non-fundamental levels of reality,
such as the level at which phenomenal properties are instantiated (if physicalism is true), which are
grounded in the fundamental level. Levine then denies that the description of the fundamental level a
priori entails descriptions of the other levels. But there is no need to postulate levels of reality.
Grounding is not a relation between levels of reality at all, because there are not multiple levels. There
is just one level of reality consisting in reality herself. But there are different descriptions of reality,
and some descriptions are more detailed or informative than others. One can still state true non-
detailed descriptions of reality provided that such descriptions are a priori entailed by the more
detailed descriptions. And when we reductively explain some phenomenon in terms of some “lower-
level” phenomenon, we are really just relating less-detailed descriptions of some aspect of reality with
more detailed aspects of that very same aspect of reality. We can describe the contents of a region of
space in terms of a composite body with certain mass—leaving out exact microphysical details, or we
can describe the contents of that region in terms of many elementary masses. The choice of detail is
up to us, it does not flow somehow from a “layered” structure of reality.

We should only postulate a distinction between metaphysical fundamentality and conceptual
fundamentality, if reality is layered into levels—some more fundamental than others. For then the
relation between these levels is the natural referent of ‘grounding’ and the bottom level that grounds
the rest is the natural referent of ‘fundamental’. Conceptual fundamentality then refers to something
different: an a priori entailment base of propositions. But without levels of reality, the natural referent
of ‘grounding’ becomes a relation among propositions (describing real properties) and the natural
referent of ’fundamental’ becomes the propositions (describing real properties) in the minimal a priori
entailment base. There seems to me to be no good reason to postulate levels of reality. And if the only
way to save physicalism against the AET/RET arguments is to postulate all these levels, then
physicalism should be abandoned.
4.1 Measurement and Reality

Using quantum worlds (worlds in which quantum mechanics is true) as test cases for AET is problematic. This is because of widespread disagreement about what is fundamental in quantum worlds and whether quantum mechanics is even capable of giving a complete fundamental description of a physical system. However, if we have independent reason to believe AET then we should be able to use AET to help resolve some of this disagreement. In chapter two I argued that AET is supported by Newtonian possible worlds, and I provided the resources for demonstrating the same for relativistic worlds. In chapter three I responded to a variety of challenges to AET and showed how certain higher level reductions support AET. I will now assume that AET is true and apply it to debates about what is fundamental and what is grounded, in quantum worlds.

I assume that quantum mechanics aims to give a complete fundamental description of reality (or at least, of some limited domain of interest e.g. physical systems not affected by gravity). Putting this assumption together with the assumption that AET is true suggests a research program. In particular the conjunction of these assumptions entail that a satisfactory quantum theory must provide the resources for a description of reality that a priori entails all truths. The research program, then, looks at proposed quantum theories and evaluates them against this constraint, perhaps with the aid of a priori entailment expansions. This chapter is intended to make some preliminary advances in this direction. In particular, I will use AET to argue that a particular class of quantum theories—dynamical collapse theories—are false.

Before delving into the quantum formalism, one can immediately state, in terms of AET, why quantum metaphysics is so difficult, and why there is so much disagreement over what’s fundamental in quantum worlds. If one attempts to evaluate quantum theory with a priori entailment expansions, one will likely find oneself getting stuck at the very first expansion conditional. That is, one will find that a quantum mechanical version of [Composition] is difficult to come by. In the classical case, we have an unambiguous Euclidean (or Riemannian) space within which little bits of mass are distributed. And so just like the (Dots-to-Shapes) expansion (section 1.5), we can apply our concept of
a composite, by grouping these masses, naming them e.g. ‘composite such and such’, and then inferring composites, “composite such and such exists”. We can then track their trajectories and observe their responses to forces in nomological possibilities, and then infer all sorts of interesting facts about those composites. Things just aren’t that easy in quantum mechanics. First of all, we don’t have an unambiguously defined “space” in which the fundamental objects are distributed. I will discuss this in detail in section 4.2. Secondly, even if we take a stand on the space in which quantum objects are distributed, what those objects actually are, and how to use the quantum formalism to understand how they are distributed in space is very tricky. I will discuss this in detail in section 4.3 in the context of dynamical collapse theories.

I understand quantum theories as what are sometimes called interpretations of quantum mechanics, and will use these notions interchangeably. Thus, quantum physicists invented a formalism to describe measurements in which quantum behaviour is apparent. Extending that formalism to a complete understandable theory—interpreting it—is very difficult. One difficulty arises from what is often called the measurement problem. Following other authors I find it useful to also distinguish a related problem: the reality problem. Quantum theories as I understand them here are attempts to solve the measurement and reality problems, within a complete fundamental theory, whose formalism stays as close as possible to the formalism developed by quantum physicists. In the remainder of this section I introduce some of the basic formalism (which I build on throughout the chapter) and I define the reality and measurement problems, before summarising what is to come.66

The state of a physical system in quantum mechanics is given by the quantum state of its wavefunction Ψ. The quantum state of a particle located at \((0_x,0_y,2_z)\) is, in Dirac notation:

\[
Ψ_1 = |(0_x,0_y,2_z)> 
\]

The symbols |> indicate a vector, and so here, the state Ψ_1 is described by a single vector in a threedimensional space pointing in the z-direction. Intuitive physical states like this are rare, however. Physical systems are typically in superpositions of such states, for example, the state of being in a superposition of being located at \((0_x,0_y,2_z)\) and being located at \((6_x,0_y,0_z)\) is represented as follows:

\[
Ψ_2 = \frac{1}{\sqrt{2}} |(0_x,0_y,2_z)> + \frac{1}{\sqrt{2}} |(6_x,0_y,0_z)> 
\]

66 I follow much of the literature and keep the discussion within the confines of non-relativistic quantum mechanics. This is usually justified by appeal to the claim that the reality and measurement problems arise in relativistic quantum field theory and even quantum gravity theory, such that solutions to such problems in quantum mechanics will yield solutions within the latter theories. See for example Butterfield (2001).
The formalism describes the particle in terms of a sum of two weighted vectors in a three dimensional space, the first pointing in the z-direction, the second pointing in the x-direction. Thus, quantum states are typically weighted sums of intuitive physical states, that is, superpositions of such states. The component states are weighted by what are called amplitudes, which are typically complex numbers. We can determine the probability that a position measurement will find the system in one or the other of its component states, with the Born Rule. The Born rule calculates the modulus square of the amplitudes. Thus, if we are to measure the position of a system in state $\Psi_2$, there is a 0.5 chance we will find it at (0, 0, 2) and there is a 0.5 chance that we will find it at (6, 0, 0). Superpositions are unintuitive physical states. With this in mind, we can define the reality problem:

**The Reality Problem**: When the physical state of a system is fundamentally described in terms of a superposition, what is the metaphysical nature of the system and how (if at all) can it ground non-fundamental truths?\(^{67}\)

Note that this problem is intimately connected to the difficulty in formulating a [Composition] expansion conditional in the quantum context.

The Born Rule allows us to get experimentally verifiable results. But what happens to our physical system (whatever it is) when we measure it? Its superposition state appears to randomly “collapse” into a definite state, such as state $\Psi_1$, upon measurement. But our act of measuring systems is surely not a fundamental process that collapses superposed states into intuitive physical states. So by what mechanism does the system move from a state like $\Psi_2$ into a state like $\Psi_1$? This is the measurement problem:

**The Measurement Problem**: Why does the act of measuring a physical system appear to collapse superposition states into intuitive states, and how can we describe what is actually happening here in fundamental physical terms?\(^{68}\)

Solutions to the measurement problem are often categorised into three types: dynamical collapse theories, hidden variable theories, and Everettian or many worlds theories. I will mostly be discussing dynamical collapse theories, although I will provide a general discussion of quantum theories in the final section.

Collapse theories aim to solve the measurement problem by specifying a collapse mechanism in unambiguous physical terms. The most well known such theory—the GRW theory\(^ {69}\)—postulates

---

\(^{67}\) The name ‘Reality Problem’ comes from Pearle (2009: 258): “There is no well-defined procedure within standard quantum theory for, at any time, plucking out from the state vector the possible states which describe what we see around us”. See also Schlosshauer (2011: 153).

\(^{68}\) For discussion see Maudlin (1995). What I call the measurement problem roughly corresponds to what Maudlin calls the problem of outcomes. See also Albert (1994: 73-79).
spontaneous collapses. In particular, wavefunctions of elementary particles spontaneously collapse, with a certain probability per unit time. The probability is extremely low, so that isolated particles rarely collapse. In contrast, interacting particle clusters, and therefore the systems they compose, are effectively in a constant state of collapse, thanks to the phenomenon of entanglement. If the states of two particles are entangled, then if one collapses, the other does too. This means that the probability of a macroscopic system (with interacting parts) collapsing at any one time is extremely high. This solution to the measurement problem is also intended to partly solve the reality problem: non-fundamental truths are grounded in collapsed superpositions. So the metaphysical obscurity of non-collapsed superpositions is no bar to understanding the manifest world in quantum terms. I will explain entanglement in more detail in the next section, and the GRW theory in more detail in section 4.3.

4.2 Dimensionality

There is widespread disagreement over the dimensionality of space according to quantum mechanics: is it three-dimensional or is it 3N-dimensional where N is the number of elementary particles? Accordingly, formulating a [Composition] expansion conditional in a quantum a priori entailment expansion is problematic. Recall the standard form of [Composition] expansion conditionals:

\[
\text{[Composition]} \quad \text{If [Fundamental microphysical description] and [T] then there is a composite C composed of } o_1 \text{ and } o_2, \text{ that is located where } o_1 \text{ and } o_2 \text{ are located (the set of points } \{x_1, x_2\}), \text{ in virtue of } o_1 \text{ being at } x_1 \text{ and } o_2 \text{ being at } x_2. 
\]

Here o is any kind of possibly fundamental object (such as a point mass, point charge, or “quantum object”). In the quantum context, any unclarity in the spatial dimensionality of o_1 and o_2 is inherited by C. And insofar as C’s relationship to space is unclear, then C’s behaviour in space will be too, and this unclarity may prevent any other expansion conditionals from being formulated. Thus, if AET is to be applied to quantum mechanics, this unclarity needs to be removed.

---


70 Although it is engrained in the literature, many prefer not to use particle-talk here. One reason is that in quantum field theory, the number of particles in a region becomes frame-variant, which would make N, and hence the dimensionality of space, impossible to define in fundamental terms (Wallace and Timpson 2010: 705). Thus as Ney (2012: 537) puts it, “the dimensionality of configuration space depends in a basic way only on the number of independent variables required to completely specify the state of the wavefunction at a time, not on anything having to do with particles in a three-dimensional space. This particle characterization of configuration space is useful as a heuristic, as it will allow us to hone in on approximately the right number of dimensions in our universe’s configuration space, but it is not strictly speaking correct”.

120
Three primary solutions have been defended in the literature. A popular view treats fundamental quantum objects as distributions in a 3N-dimensional space. Another view expands the fundamental ontology to include three dimensional objects that bear fundamental nomic relations to the fundamental 3N-dimensional objects. The third view rejects the 3ND ontology in favour of a purely three dimensional ontology.

In my view, advocates (and critics) of the 3ND view are not clear on precisely what the argument for the 3ND view is. Consider Bell’s oft cited claim:

“There is nothing in this theory but the wave-function. It is in the wave-function that we must find an image of the physical world, and in particular of the arrangement of things in ordinary three-dimensional space. But the wave function as a whole lives in a much bigger space (than physical space), of 3N dimensions. It makes no sense to ask for the amplitude or phase or whatever of the wavefunction at a point of ordinary space. It has neither amplitude nor phase nor anything else until a multitude of points in ordinary three space are specified”.

(Bell 1987: 204)

Many authors quote this claim and simply take it for granted and assume the 3ND view. But the claim can be challenged. Consider $\Psi_2$ from the previous section. Clearly we can ask after the amplitude at a point in ordinary space. Of course Bell and others are more concerned with entangled states, where amplitudes become associated, not with points in space, but with entire particle configurations. But the key question is: why should that motivate an increase in spatial dimensions in fundamental ontology? To better understand this, we require a better understanding of entanglement.

Let’s say we have a particle $p_1$, located at $(x, y, z)$ at $t_1$, which is travelling towards location $(0, y, 2z)$. And let’s say there is just one other stationary particle $p_2$ and that their interactions obey the following rules:

If $p_2$ is located at $(0, 0, 2z)$, then at $t_2$, $p_1$ will be located at $(1, 0, 1z)$.

If $p_2$ is not located at $(0, 0, 2z)$, but rather at $(6, 0, 0z)$, then at $t_2$, $p_1$ will be located at $(-1, 0, 3z)$.

This means that if at $t_1$, $p_1$ co-exists only with a stationary particle $p_2$, whose state is given by $\Psi_2$, then at $t_2$, $p_1$ will have evolved into an entangled superposition with $p_2$ so that the two particles can only be jointly described as the entangled quantum state $\Psi_3$:

71 For example, Allori et. al. (2008: section 4.3). Similarly Maudlin (2007: 3165), who only adds that “if we fire entangled pairs of particles at the slits, the wavefunction for the pair is no longer a function on something isomorphic to physical space: it is a function on a space of twice as many dimensions”.

121
ψ₃ = \sqrt{\frac{1}{2}} [ |(1_x|0_y|1_z) > + |(0_x|0_y|2_z) > ] + \sqrt{\frac{1}{2}} [ |(-1_x|0_y|3_z) > + |(6_x|0_y|0_z) > ]

This is the wavefunction for the now entangled two-particle system. An implication of this entanglement relation is that if p₁ collapses into location \(1_x|0_y|1_z\), then p₂ will immediately collapse into \(0_x|0_y|2_z\). Similarly, if p₂ collapses into \(6_x|0_y|0_z\), then p₁ will immediately collapse into \((-1_x|0_y|3_z\), and so on. This is entanglement.

With this in mind, why would the above state motivate the idea that space is 3N-dimensional (in this case, 6-dimensional), particularly when, at least on the face of it, the only dimensions explicitly mentioned (four times, as it happens) by \(Ψ₃\) are the x, y and z dimensions? Neither p₁ nor p₂ has a definite location in the intuitive sense; they are instead both in superpositions of different positions. But is there any natural way to describe these weighted sums of intuitive position states as states in a three dimensional space? If not then we have an argument from entanglement, for the 3ND view.

The obvious way to determine this is to draw up three dimensional graphs, to see if we can capture, in these graphs, all of the information that \(Ψ₃\) contains. The idea is that doing so allows us to clearly conceive of quantum mechanical properties in terms of distributions of properties in three-dimensional space. And if at some point we come across a quantum mechanical property that we cannot capture three-dimensionally, then this is evidence that we cannot think of wavefunctions as evolving in a three dimensional space.\(^{72}\)

Thus, we might draw up our (x, y, z) dimensions, and start by placing dots at each point that exemplifies non-zero amplitude, which in the case of \(Ψ₃\) are the points \(1_x|0_y|1_z\), \(0_x|0_y|2_z\), \((-1_x|0_y|3_z\) and \(6_x|0_y|0_z\). We then need a way of distinguishing our two particles. To do this, we can replace the dots at \(0_x|0_y|2_z\) and \(6_x|0_y|0_z\) with little squares: dots for p₁ superposition components, squares for p₂ superposition components. We also need to distinguish distinct amplitudes. Perhaps we can associate a given amplitude with a particular colour. Let’s associate the square root of 0.5 with white. In that case, our graph will contain two white dots, one at \(1_x|0_y|1_z\) another at \((-1_x|0_y|3_z\), and it will contain two white squares, one at \(0_x|0_y|2_z\) another at \(6_x|0_y|0_z\). The resulting graph might look something like this:

---

\(^{72}\) My formulation of the problem follows that of Ney (2012), and to some extent Lewis (2004).
Have we accounted for all the information in $\Psi_3$? If we have, there is no problem: all relevant information can be accounted for in a three dimensional space, and so “recovering” a three dimensional space from the fundamental description is trivial: we just treat the three dimensional representation as fundamental, just as in classical mechanics.

However our graph leaves out crucial information contained in $\Psi_3$. In particular, we have left out the information that physically distinguishes $\Psi_3$ from $\Psi_4$:

$$\Psi_4 = \frac{1}{\sqrt{2}} \left[ |(1_x|0_y|1_z) > + |(0_x|0_y|2_z) > \right] + \frac{1}{\sqrt{2}} \left[ |(-1_x|0_y|3_z) > + |(0_x|0_y|2_z) > \right]$$

Now, if we draw up a three-dimensional graph for $\Psi_4$ using white dots and white squares, the graph will be identical to the one above. But $\Psi_4$ is not identical to $\Psi_3$, which suggests that three-dimensional graphs are insufficient to capture the states of certain quantum systems.

What exactly is the information that our three-dimensional graphs are missing? In the two equations, all we have done is swap the $(6_x|0_y|0_z)$ with the $(0_x|0_y|2_z)$, that is, we have swapped the two superposition components of $p_2$. This has drastic physical effects. For example, if our system’s state is $\Psi_3$ and $p_1$ collapses into $|(1_x|0_y|1_z) >$, then $p_2$ will collapse into $(0_x|0_y|2_z)$; but if our system’s state is $\Psi_4$ and $p_1$ collapses into $|(1_x|0_y|1_z) >$, then $p_2$ will collapse into $(6_x|0_y|0_z)$.

Can we capture this entanglement information in our three dimensional graph? Typical presentations of the problem assert that we cannot, for example, Ney states that:

“If we want an adequate characterization of either state, one that distinguishes $\Psi_3$ from $\Psi_4$, we will need to move to a higher dimensional configuration space.”

(Ney 2012: 30-31)
Similarly Lewis states that:

“Unlike the classical case, we cannot regard the configuration space representation merely as a convenient summary of the individual particle states; there are physical properties of the two-particle system [entanglements] that are only captured in the configuration space representation. [...] The inescapable conclusion for the wavefunction realist seems to be that the world has 3N dimensions.”

(Lewis 2004: 717)

When physicists capture this information, they invoke configuration spaces. They are 3N-dimensional, where N is the number of particles in the system. The problem is how to explain the three dimensions of our experience in terms of the 3N dimensional configuration space. And this certainly looks like a daunting problem. Monton has put the point in a particularly precise manner, by arguing that three-dimensional particles cannot supervene on any 3ND wave function:

“The reason there is no supervenience is that nowhere in the 3N-dimensional space is it specified which dimensions correspond to which particles. It could be that the x, y, and z coordinates of particle number 3 in the three-dimensional space correspond to the seventh, eighth, and ninth dimensions of the 3N-dimensional space, or it could be that they correspond to the ninth, eighth, and seventh dimensions [...] and so on. These different correspondences entail different ways that the N particles evolve, given a particular evolution of the objects in the 3N-dimensional space [...] given the state of the objects in 3N-dimensional space, one cannot establish the state of the objects in three-dimensional space.”


Another way to put Monton’s point is that one cannot establish $\Psi_3$ from above, given the following quantum state:

$$\Psi_{3*} = \sqrt{\frac{1}{2}} \left[ \left| (1_a|0_b|1_c|0_d|0_e|2_f) > \right| + \sqrt{\frac{1}{2}} \left| (-1_a|0_b|3_c|6_d|0_e|0_f) > \right| \right]$$

The two noteworthy features of $\Psi_{3*}$ are that the (x,y,z) coordinates have been replaced with 3N distinct coordinates (a-f) and the particle labels (1,2) are gone from the superposition components: there is now just a single system, with amplitudes distributed over two distinct points in a six dimensional space. The ontology of the 3ND view consists in a single object—a wave function—which (in our simple example) is a function from points in configuration space to amplitudes. In the $\Psi_{3*}$ example, the wave function is a function from point (1,0,1,0,0,2) to a particular amplitude and from point (-1,0,3,6,0,0) to the same amplitude. Monton’s point is that we can treat the a-dimension and the d-dimension as the x-dimension if we wish, but the fundamental description does not force
this upon us. For this reason, the 3ND wave function description does not a priori entail three dimensional particles.

Albert goes so far as to suggest that our experience of space is completely illusory:

“[I]t has been essential to the project of quantum-mechanical realism [...] to learn to think of wave functions as physical objects in and of themselves. And of course the space those sorts of objects live in, and (therefore) the space we live in [...] is configuration-space. And whatever impression we have to the contrary (whatever impression we have, say, of living in a three-dimensional space, or in a four-dimensional space-time) is somehow flatly illusory”.

(Albert 1996: 277)

But let’s back up. In response to Ney’s claim that we need to move to a higher dimensional space, why can’t one just add more structure to our three dimensional graph to distinguish $\Psi_3$ from $\Psi_4$, why must we move to the configuration space, in the context of doing quantum metaphysics?

As noted, the argument for the 3ND view is seldom spelt out, which is why the argument of Ney (2012: appendix), summarised above, is an important contribution to the literature. But I think the style of argument also gives the game away. Recall that in building our graph, we represented particles in terms of shapes (dots and squares) at each point in the graph that those particles have superposition components with non-zero amplitude. We also represented the amplitude values with colours. The problem is then how we represent entanglement relations. One option is to move to a configuration space representation, thereby increasing the number of dimensions. Now, while this may be convenient in the context of doing physics, it is unnecessary when doing quantum metaphysics. To represent entanglement relations on the graph, just add further symbols! Thus, to distinguish $\Psi_3$ from $\Psi_4$, I have added a star to the white dot at (1,0,1), a star to the white square at (0,0,2), a sun to the white dot at (-1,0,3), and a sun to the white square at (6,0,0):

---

73 Lewis (2004: 715-6) also attempted to flesh out an argument for the 3ND view, by appeal to what can be represented in three dimensional graphs. However, he argued that if two particles are entangled, then one cannot represent the two particles separately, in their own separate three dimensional graphs. This might be right but it is beside the point. What’s relevant is whether we can represent the entangled wavefunction of the two-particle system in a single three dimensional graph. Lewis’ argument does not speak to this, whereas Ney’s (2012: appendix) argument is an attempt to.
So I am essentially saying that the configuration space is as necessary to quantum mechanics as it is to classical mechanics. And just as in classical mechanics we treated the configuration space representation as a useful fiction, instrumental in describing what is really a three dimensional space, so too should we say the same in quantum mechanics. Why ontologically privilege the three dimensional representation in quantum mechanics? For the same reason we always did in classical mechanics: let’s believe what we see until we have some reason to think otherwise.

One reason to be sceptical of this 3D view, which treats entanglement as a fundamental relation, concerns what happens when we consider more complex systems. Consider what happens when we move from a simple two particle entanglement, to macroscopic bodies. And let’s for the moment work with an Everettian theory: i.e. one that does not alter the usual dynamics (given by the Schrödinger
equation) to produce wavefunction collapse and does not add additional ontology to the existing ontology (the wavefunction). According to the usual dynamics, if the state of a macroscopic system is correlated to a microscopic superposition state, then that superposition state is magnified up into the macroscopic realm. For example, consider a measuring device that is set up to detect the position of a particle, so that the pointer on the measuring device points to the left if it finds the particle at (0,0,2), and points to the right if it finds the particle at (6,0,0). Now imagine we use this device to measure the position of the particle in state $\Psi_2$ (a superposition of (0,0,2), and (6,0,0)). Then according to the usual dynamics, the measuring device will evolve into a superposition of pointing left and pointing right. So where $M$ represents the measuring device, and $p$ represents the particle, we have the following entangled quantum state:

$$\Psi_5 = \frac{1}{\sqrt{2}} \left[ |(\text{LEFT}) >_M |(6_x|0_y|0_z) >_p \right] + \frac{1}{\sqrt{2}} \left[ |(\text{RIGHT}) >_M |(0_x|0_y|2_z) >_p \right]$$

Now consider a 3D graph representing this state, and ask where each of the $M$ superposition components are located. They are of course overlapping each other (except for the top parts of the pointers, which point in different directions). And it is not just measuring instruments (or their superposition components) that overlap, so do people. Thus, if an experimenter has correlated the state of her brain to the state of the measuring instrument (i.e. is ready to become conscious of the resulting pointer state), then according to the standard dynamics, she will evolve into a superposition of experiencing a left result, and experiencing a right result. But the positions of her two superposition components do not move on this 3D view (except the neurons etc. corresponding to the differing experiences). I suspect that this is what is truly troubling theorists about a three dimensional quantum mechanics: it isn’t entanglement, so much as a fear of overlap.

The usual Everettian analysis of the above measurement scenario is in terms of many worlds. Thus, when the measurement magnifies the microscopic superposition up into the macroscopic realm, the measuring device splits into two different measuring devices (one pointing left, the other pointing right). And since the experimenter, as well as the rest of the universe, is entangled with the measuring device, the entire universe bifurcates into two universes. On one view, all of space bifurcates too, so that each world evolves in its own three dimensions, (or into its own subspace within configuration space). On the present view, space does not bifurcate in any sense, only the matter within it does.

I have explained how 3D view yields overlap in the Everettian or many worlds theory. But overlap occurs on any quantum theory. The standard additional variables theory (Bohmian mechanics) does not reduce the wave function ontology, but only adds to it. In particular, special bits of matter—corpuscles—are added to one of the superposition components while the rest are treated more like fields of influence. So these fields must overlap without actually interacting. And even for dynamical
collapse theories, overlap is inevitable. As I discuss in detail in the next section, dynamical collapse does not destroy all the superposition components but one—they are still there, and whatever they are, they spatially overlap the collapsed components.

I see no overlap-based objection to the 3D view. For while the overlapping entities are not interacting this is not some mysterious phenomenon. It is instead a straightforward consequence of treating entanglement (correlation) as a fundamental physical relation. This point is recognised by Lewis:

“[T]wo wave packets that are components of the state of one and the same particle sometimes interact and sometimes pass by each other when their three-dimensional coordinates coincide. Doesn’t this require the existence of extra dimensions in which the passing-by can take place? Certainly one needs parameters in the theory, the values of which determine whether or not the packets interact. And in the quantum case, the parameters in question refer to the coordinates of the other particles in the system—i.e. they encode how the wave packet for the particle we are following is correlated with the wave packets for the other particles. But as argued above, the structure underlying such correlations need not be regarded as itself spatial.”

(Lewis 2013: 124)

So other than fear of overlap, there appears to be no reason to reject the 3D view and opt for the 3ND view. And there is reason to reject the 3ND view: it apparently fails to a priori entail the manifest 3D world.74

There are further problems with the 3ND view. Consider the following question: why is it that in every nomologically possible world (worlds with the same laws but different initial conditions), the dimensions of the universe are a multiple of three? On the 3D view, the answer is trivial: because the theory only posits three fundamental spatial dimensions. But notice how awkward this question is for those who take the 3ND configuration space to be fundamental. The view allows for a 12 dimensional world and a 15 dimensional world, but rules out a 13 dimensional world. Why?

Ney (2012: 14) admits that this question cannot be answered when she states that “the dimensionality of the space the wavefunction inhabits is not determined by anything more fundamental”. But this sort of structure is not something to be taken as brute as it involves an intimate relationship between the structure of space and the stuff that evolves within it. In what does this relationship consist? The 3D view posits no such dependence and so has no need to explain it.

74 It should be noted that this is certainly not obvious. For example Albert (2013) argues that the diachronic behavior of a wave function in 3ND space may “functionally enact” the existence of tables and chairs. The main focus here is to argue that the argument against the 3D view, which motivates the 3ND view, does not work.
The resulting view is similar to Monton’s (2002, 2006, 2013) view and Lewis’s (2013) considered view. Monton describes the wavefunction as a holistic fundamental property exemplified by multiple particles in 3D space. Lewis (2004, 2013) notes that while the wavefunction is 3ND in the sense that one requires 3N parameters to specify the physical properties of a given state, that is quite different from saying that the wavefunction evolves in any more than three spatial directions.

The primary point that I want to make is that if we are constructing possible worlds then we can quite easily construct the sort of world that Lewis and Monton envisage. We can also construct a possible world in which space has 3N dimensions (in the directional sense) and where three-space is an illusion for the creatures living within it—so we can easily construct the world that Albert envisages. The question is then simply which possible world is actual. Or better, since we are concerned with non-relativistic quantum mechanics: which possible world is closest to the actual world. Here, it is crucial to note that we see a three-dimensional space. Of course this is consistent with Albert’s view that postulates such seeings, but treats them as non-veridical. But here I side with Monton (2006), who argues that the 3D view should be favoured given that it is a far more conservative view than the 3ND view, and this certainly seems right given that on the 3D view we don’t have to treat macroscopic space as illusionary. But it’s worth noting, and this will be important later, that the evidence does not appear to decide between either view, which is precisely principles like conservatism are what we must appeal to, to select our favoured ontology.

We are one step closer to being able to formulate a quantum version of expansion conditional [Composition]. In particular, when we consider a world described by quantum theory, we can treat the space in which we infer non-fundamental composites as three-dimensional, and so we can treat our non-fundamental composites as 3D. However, it is still not clear what sort of entities we should group together in thought, name, and treat as non-fundamental composites. This brings us to the tails problem for dynamical collapse theories.

### 4.3 Dynamical Collapse Theories

Let’s return to physical state $\Psi_2$ and note that we are thinking of it as an amplitude distribution over three-dimensional space:

$$\Psi_2 = \frac{1}{\sqrt{2}} |0_x|0_y|2_z\rangle + \frac{1}{\sqrt{2}} |6_x|0_y|0_z\rangle$$

129
This is an equal superposition, in the sense that the component amplitudes are equally distributed over two distinct locations \( ((0_x|0_y|2_z) \) and \( (6_x|0_y|0_z) \). The GRW collapse dynamics is meant to help by “collapsing” the superposition into one of its components, either \( (0_x|0_y|2_z) \) or \( (6_x|0_y|0_z) \). But this is metaphorical. What the collapse literally does is shift the amplitudes around. However, the collapse does not put all the total amplitude in one localised spot (e.g. \( (0_x|0_y|2_z) \)). Rather, if the particle spontaneously collapses, it will transform into something like this (where \( \alpha \) is very small, but non-zero):

\[
\Psi_6 = \sqrt{1 - \alpha^2} |(0_x|0_y|2_z) > + \alpha |(6_x|0_y|0_z) >
\]

If the collapse localised position amplitudes to specific regions then certain unwanted violations of energy conservation would occur. By the uncertainty principle, the more concentrated a wave function’s amplitude for position is, the more dispersed its amplitude for momentum. But the more dispersed momentum amplitude is the more energy the system can possess. So to prevent systems from spontaneously heating up in ways we don’t observe, the collapse must disperse position amplitude somewhat.\(^{75}\) Thus, something still appears to exist at \( (6_x|0_y|0_z) \): something which exemplifies amplitude \( \alpha \). Furthermore, the precise degree of amplitude distribution at \( (6_x|0_y|0_z) \), as small as it is, has a real effect on the overall dynamics of the system.

If one measures the position of \( \Psi_2 \) one will entangle \( \Psi_2 \) with all the particles that make up one’s measuring instrument and one’s body. As these particles are themselves all entangled with each other already, their probability per unit time for collapse is extremely high. Hence, in measuring \( \Psi_2 \), we collapse it, most likely to state \( (0_x|0_y|2_z) \). The same basic idea still applies: \( (6_x|0_y|0_z) \) will still have (a shared) amplitude dispersed over it. So it is not at all clear that the collapse localises anything: it just shifts amplitudes around. There is still some physical stuff located at \( (6_x|0_y|0_z) \), just as there was prior to collapse. The problem applies at all scales. The wavefunction for a table (the entangled wavefunction of all its parts) is smeared out all over space. Why does the fact that it has high amplitude within the region I take to be my office, suggest that that is where my table is located? And so we have the tails problem:

**Tails problem:** it is unclear how the collapse mechanism can solve the measurement problem, just by changing amplitude values at points in space, without reducing them to zero.\(^{76}\)

---

\(^{75}\) See Clifton and Monton (1999: 698). Lewis (1995: 29) notes that representing collapse by multiplying the wave function of a particle by a step function, rather than a Gaussian, will not help. Not only will collapse only confine such particles for an instant of time, but because realistic entanglement is never perfect, the system entangled with the collapsed particle will not collapse to a definite finite region.

\(^{76}\) See Lewis (1995) for detailed discussion.
A number of solutions to the tails problem have been proposed in the literature. There is a striking parallel between these solutions and physicalist solutions to the mind-body problem discussed in section 3.4. The following table demonstrates the parallels, and will help structure the discussion:

<table>
<thead>
<tr>
<th><strong>A Posteriori Reductionism:</strong></th>
<th><strong>A Posteriori Physicalism:</strong></th>
<th><strong>A Posteriori Wavefunctionism:</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Reject AET/RET.</td>
<td>Phenomenal truths are grounded in physical truths, but do not follow a priori from physical truths.</td>
<td>Particle truths are grounded in wave function truths, but do not follow a priori from wave function truths.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>A Priori Reductionism:</strong></th>
<th><strong>A Priori Physicalism:</strong></th>
<th><strong>A Priori Wavefunctionism:</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Accept AET/RET.</td>
<td>Phenomenal truths are grounded in physical truths and follow a priori from physical truths.</td>
<td>Particle truths are grounded in wave function truths and follow a priori from wave function truths.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Eliminativism:</strong></th>
<th><strong>Eliminativist Physicalism:</strong></th>
<th><strong>Eliminativist Wavefunctionism:</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Explanandum is an illusion.</td>
<td>There are no phenomenal properties—physical truths are fundamental and consciousness is an illusion.</td>
<td>There are no particles—wave function truths are fundamental and particles are an illusion.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Dualism:</strong></th>
<th><strong>Epiphenomenal Dualism:</strong></th>
<th><strong>Wave Particle Dualism:</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Explanandum is irreducibly fundamental.</td>
<td>Phenomenal properties are causally inert fundamental properties, caused by fundamental physical properties.</td>
<td>Particles are causally inert fundamental entities, caused by fundamental wave function properties.</td>
</tr>
</tbody>
</table>

Just as physicalism asserts that only physical truths are fundamental, wavefunctionism asserts that only wavefunction truths are fundamental. And like we see in the physicalism literature, a wavefunctionist might accept AET/RET and therefore require that all truths follow a priori from wave function truths, or they might not. Wavefunctionists might also deny the very existence of the explanandum, in the way that eliminativist physicalists deny consciousness. Finally, one might hold that although the wavefunction is causally closed, further fundamental entities—localised particles—are caused to exist by the wavefunction in accordance with further fundamental laws. In what follows
I use AET to evaluate these positions, and find them all to be implausible. This will illustrate the usefulness of AET in debates in quantum metaphysics.

**Against A Posteriori Wavefunctionism**

In their influential paper ‘Tails of Schrödinger’s Cat’ (1996), David Albert and Barry Loewer try to solve the tails problem by proposing a bridge principle linking wave function descriptions to descriptions of localised particles. This quantum to classical bridge principle has come to be known as ‘the fuzzy-link’. Albert and Loewer state the principle as follows:

“‘Particle x is in region R’ iff the proportion of the total mod-square value of x’s quantum state associated with points in R is greater than or equal to 1-p’.

(Albert and Loewer 1996: 87)

The value ‘p’ is left vague, but is intended to be close to zero. Albert and Loewer then add that “[The fuzzy link principle] is a new proposal, an alternative proposal, about the relation between position-talk and quantum talk; a new proposed supervenience rule” (1996: 87).

Albert and Loewer simply *postulate* this principle rather than explain it in terms of the wave function. This is why I classify their view as an *a posteriori wavefunctionist* collapse theory. Thus, their view is analogous to a posteriori physicalists who postulate psychophysical bridge principles to solve the mind-body problem, while denying the need to explain such principles in physical terms.

If the fuzzy link principle were a priori, or a priori deducible from the quantum state, then it could be used in defence of a quantum [Composition] expansion conditional. However, the principle is neither a priori nor a priori deducible from the quantum state. This can be seen in the following symmetry argument. Recall quantum state $\Psi_5$ in which the measuring device's pointer state becomes entangled with the position of the particle being measured. According to collapse theories, this kind of large scale entanglement will induce collapse. So if the particle collapses into position $(6_x|0_y|0_z)$ then the quantum state is something like this:

$$\Psi_7 = \sqrt{1-\alpha^2} \left[ |(LEFT) >_M (6_x|0_y|0_z) >_p \right] + \alpha \left[ |(RIGHT) >_M (0_x|0_y|2_z) >_p \right]$$

Albert and Loewer want to say that $\Psi_7$ grounds a situation in which the device points to the left and the particle is located at $(6_x|0_y|0_z)$. This is because the modulus square of $\sqrt{1-\alpha^2}$ is close to one whereas the modulus square of $\alpha$ is miniscule. The problem is that the structural symmetry between

---

77 The label comes from Clifton and Monton (1999: 699).
78 Lewis (2006: 232) describes the principle as “a proposal for how to use language”.
the two superposition components is such that if we can infer the existence of the left component we should also be able to infer the existence of the right component too. For nothing in this description ontologically favours one state over the other and merely having a higher mod square value is insufficient to break the structural symmetry. A similar worry is expressed by Wallace, who uses the example of Schrödinger’s cat:79

“If the link principles are just a matter of descriptive convenience then what prevents us regarding observers as being just as present in the dead-cat term as in the live-cat term? After all [...] the dead-cat term is as rich in complex structure as the live-cat term”.  

(Wallace 2008: 63)

I take Wallace's point to be this: if Albert and Loewer can just freely postulate a bridge principle that ontologically privileges high mod-square components, then why can't we similarly postulate a bridge principle that privileges low mod-square components. This points to a more general problem with a posteriori views of grounding more generally: when the constraints on grounding (or supervenience) claims become so weak, grounding claims start to look ad hoc, particularly when they are proposed solely because, were they true, they would solve the problem. AET allows us to rule out such ad hoc hypotheses.

Interestingly, Albert and Loewer suggest that their proposal is not different from what we are used to in classical physics.80 This is in direct conflict to what I have been arguing: that classical physics exhibits a priori entailment while Albert and Loewer's collapse theory does not. They claim that:

“It was one of the upshots of classical mechanics, after all, that our everyday macroscopic physical talk, our talk of tables and chairs and baseballs and pointers on measuring-devices, supervenes only vaguely on the exact physical micro-language. All we’re being confronted with here is that representing the world entirely by means of quantum-mechanical wave-functions is going to require that we swallow one additional level of vagueness: that our everyday language will supervene only vaguely (just as it always has) on the micro-language of particle positions, and that that language will itself supervene only vaguely [...] on the fundamental language of physics”.

(Albert and Loewer 1996: 91)

The problem with this argument is that the micro-language of particle positions does not supervene only vaguely on the fundamental physical language. It's true that our everyday language supervenes only vaguely on the micro-language of particle positions. In other words, given a microphysical description of a cat, before one can a priori infer a cat, one must a priori a large number of admissible

79 See also Cordero (1999).
80 See also Lewis (2003: 1438).
cat precisifications. In other words, from the microphysics one infers a precise set of particles that behaves like a cat (‘plays the cat role’). One can also infer another precise set containing one extra molecule, that also behaves like a cat, and so on. Eventually, one infers a number of admissible precisifications of the ordinary term ‘the cat’. This is how the vague language of cats supervenes on (is a priori deducible from) the precise language of microphysics.

Nothing like this is happening in the quantum to microphysics case. The two terms describing the two superposition components each contain propositions from which one can deduce admissible measuring device precisifications. But then we end up with two measuring devices! The problem with the quantum to microphysics case is not vagueness. The problem is that in order to deduce non-fundamental microphysical truths on Albert and Loewer’s view, we must apply a (precise) mathematical operation to the fundamental description, one which does not itself follow a priori from the fundamental description. Furthermore, this mathematical function only gives us certain numbers as outputs, and Albert and Loewer’s brute bridge principle requires that we quite arbitrarily, treat certain ‘high’ outputs as existence determining. The fact that the word ‘high’ is vague is beside the point. The same problem would apply even if ‘high’ were replaced with a precise value. Albert and Loewer’s fuzzy link formulation of GRW is inconsistent with AET and so should be rejected.

**Against A Priori Wavefunctionism**

While collapse theories are not helped by the introduction of brute bridge principles, it may be possible to formulate non-brute bridge principles that solve the tails problem. That is, it may be possible to formulate bridge principles that can be explained in terms of the quantum state and analyses of relevant notions like ‘particle x exists in region R’. Here I consider a number of options and show why they cannot work.

Recall \[ \Psi_7 \]:

\[
\Psi_7 = \sqrt{1 - \alpha^2} \left[ |(LEFT) >_M |(6x|0_y|0_z) >_P \right] + \alpha \left[ |(RIGHT) >_M |(0_x|0_y|2_z) >_P \right]
\]

Since the moduli square of the amplitudes are probabilities, perhaps it is possible to analyse existence in terms of probability. Such a view is suggested by Clifton and Monton:

“If one is willing to entertain the thought that events in a quantum world can happen without being mandated or made overwhelmingly likely by the wavefunction, then it is no longer clear why one should need to solve the measurement problem by collapsing wavefunctions! […] …one supposes there to be a plausible intuitive connection between an event’s having a high probability according to a theory, and the event actually occurring”.
In a later paper, Clifton and Monton assert that dynamical collapse theories cannot solve the measurement problem without this assumption:

“[D]ynamical reduction theories generically do not eliminate the tails of an object’s wavefunction when it collapses. Nevertheless, since \(|a|^2 \gg |b|^2\), the probability of finding any particular marble in the box is very high, and it is natural to take each individual marble to be in the box. Indeed, not doing so would leave dynamical reduction theories without a complete solution to the measurement problem”.

Clifton and Monton motivate this view with the assertion that the connection between high probability and existence is “plausible”, “intuitive”, and “natural”, and with the assertion that without this connection, dynamical collapse theories fail. However, there are two fatal problems with the view. Firstly, the connection, far from being intuitive, conflicts with the fact that existence does not come in degrees. There can be no connection between high probability and high degree of existence. So there is no natural connection between the concepts. Furthermore, the intended view is not one where existence comes in degrees, but is one where existence (simpliciter) occurs in connection with high probability. This is not at all a natural connection but appears to be more of an arbitrary stipulation. So no analysis of “a particle exists at \((x,0,0)\)” can yield a functional role played by a high probability for a particle to exist at \((x,0,0)\).

A more severe problem (noted by Wallace (2008: 59)) is that Clifton and Monton are confusing the role of probabilities in collapse theories. The mod-square of a superposition component is the objective probability for that component to become a collapse centre. In no sense is it also the objective probability for that component to have ‘actually occurring’ status, or anything of the sort. On a realist view of dynamical collapse theories, the entire wave function exists (simpliciter).

Let’s try an alternative analysis of spatial notions. One option is to analyse location phenomena in terms of the experiences they induce in observers. This option is explored by Chalmers:

“In the case of macroscopic spatial properties, it is plausible that spatial properties can be picked out by spatial concepts as that manifold of properties that serve as the causal basis for spatial experience [...] To simplify, the property of being two meters away from one might be picked out as the spatial relation that normally brings about the experience of being two meters away from one. [...] One can then argue that on a collapse interpretation, the properties and relations that normally bring about the relevant sort of spatial experiences are precisely
properties and relations requiring the wavefunction’s amplitude to be largely concentrated in a certain area.

(Chalmers 2012: 295-296)

There are two ways to make sense of this proposal, depending on whether we think the wave function is defined in a 3D space or a 3ND space. On the 3D view, it is important to collapse theories that the manifold of properties causally responsible for our experiences of being located at region R, includes the property of having high (close to 1) mod square amplitude distributed at R. On the 3ND view, it is important to collapse theories that the manifold of properties causally responsible for our experiences of being located at R includes the property of having high (close to 1) mod square amplitude distributed at a certain region in the 3ND space. Either way, if ‘being located at R’ can be analysed in terms of the causal basis for R-location experiences, then from facts about such experiences and their quantum basis we can deduce and explain location facts. (Of course, tiny particles cannot cause the relevant experiences. But the particle still exemplifies the type of property that causes the relevant experiences.) Call this the phenomenal analysis of spatial concepts.

In what follows, I argue that this analysis does not help. The crucial question is whether wave function collapse guarantees an adequate causal basis for our spatial experiences. So we can agree with the phenomenal analysis of spatial concepts but deny that collapse theories can account for spatial experiences. And we should deny this on the 3D view, based on the symmetry argument above. And we should be very sceptical of the 3ND view, due to the striking reliance on psychophysics that would be required to make the view coherent. So the phenomenal analysis does not help collapse theories.

An initial problem is that the phenomenal analysis only shows that we can deduce spatial truths from wave function truths plus phenomenal truths. But our question is whether we can deduce spatial truths from wave function truths alone. We can avoid this problem as follows: either consciousness is fundamental (dualism) or it isn’t (physicalism). If dualism is true then there is no problem, because the goal is just to derive location truths from fundamental truths, and the relevant dualism treats both the wave function and consciousness as fundamental. If (a priori) physicalism is true, then phenomenal truths are derivable from fundamental physical (wave function) truths, we just don’t yet know how. To avoid the mind-body problem, we can just add phenomenal truths to the derivability base, while staying neutral on the mind-body problem. My objection applies either way.

Let’s begin with the 3D view. Recall \( \Psi_7 \) and consider what happens, according to the collapse dynamics, to an experimenter perceiving the state of the pointer:

\[
\Psi_8 = \sqrt{1 - \alpha^2} \left[ |(PERCEIVE LEFT) >_E |(LEFT) >_M |(6x|0_y|0_z) >_p \right] \\
+ \alpha \left[ |(PERCEIVE RIGHT) >_E |(RIGHT) >_M |(0_x|0_y|2_z) >_p \right]
\]
The experimenter E has evolved into a superposition of perceiving a left result and perceiving a right result. More precisely, $\Psi_8$ is a compound state composed of two superposition components. One component describes particle configurations which entail the existence of an experimenter E perceiving a left result, and there actually being a left result (in device M). The other describes particle configurations which entail the existence of an experimenter E perceiving a right result, and there actually being a right result (in device M). These components are assigned mathematical values (amplitudes) and are related by mathematical values (e.g. $+$), which in part represent dynamical relationships.

I have used ‘perceive’ rather than ‘experience’ because I do not assume that the physical description entails anything about experiences. Here ‘perceive’ is a non-phenomenal notion which can be functionally analysed in terms of registering environmental information and making it available to other internal states (belief, memory etc.). The idea is that the proper microphysical description would describe particle configurations that play the conceptual role for perceiving.

The objection can now be stated: if we assume that perceiving a left result is always associated with experiencing a left result, then applying the phenomenal analysis entails that there is a measuring device pointing left and that there is a measuring device pointing right (and that they both overlap each other in space cf. section 4.2)). But this is the wrong result. Recall that it is important to 3D collapse theories that the properties causally responsible for our experiences of being located at region $R$, includes the property of having high (close to 1) mod square amplitude distributed at $R$. This is crucial to the empirical adequacy of collapse theories. For when we actually perform such experiments, we only experience one result. We do not bifurcate into two experimenters—that would turn the theory into an Everett interpretation, whereas collapse theories are intended to be “one world” interpretations.

The only way for collapse theories to recover the right experiences given the quantum state, is to postulate complex psychophysical laws (whether they be physicalist supervenience laws, or dualist nomological laws). The postulate would have to be that perceiving phenomenon $p$ is not sufficient for experiencing $p$: only when one perceives $p$ with high amplitude does one experience $p$. Because the collapse mechanism reduces the amplitude of one superposition component, and increases the other, this psychophysical law prevents the wrong results: while the left component is associated with an experience of an experimental result, the right component is not. But this doesn’t help, because it leaves us with a component in which a configuration of particles composes a human perceiver, perceiving an experimental result unconsciously. The view leaves behind zombie tails! Such a view is difficult to take seriously.
One way to avoid this result is to phenomenally analyse all relevant concepts including the concept of perceiving. This would allow the collapse theorist to say that there is no perceiver in the right component because nothing in that component is a causal basis of experiences of perceivers. But this view fails because not all concepts admit of phenomenal analysis. For example, in section 3.4 I discussed a class of concepts—cousin concepts—which by definition work much like phenomenally analysable concepts, except that they cannot be phenomenally analysed—they are analysed in structural-functional terms. This allows us to say that while there may be no perceivers (things that cause experiences of perceivers) in the right component, there are still perceivers* (things that register environmental information) and this is not a result that collapse theories should generate. For example, an advocate of this position might claim that one cannot even infer composites from the term for the right superposition component, because there are no causal bases for experiences of composites. But in response, we can still infer composites* (understood in accordance with universal composition), which enables us to infer that the tails are swarming with zombies*.

The general problem for 3D reductive collapse theories is that the fundamental description entails the existence of far too many entities. As the symmetry arguments show, if we can infer macroscopic entities from high amplitude components then we can infer macroscopic entities from low amplitude components too. But such entities undermine the empirical adequacy, and the one-world spirit, of collapse theories. No analyses of spatial or other concepts appear to help here, and so a priori wavefunctionist 3D collapse theories fail.

Now consider the 3ND view, and recall from the previous section our comparison of $\Psi_3$ with $\Psi_3^*$, which illustrated the difference between the 3ND view and the 3D view. When moving from the 3D representation to the 3ND representation, we replace the vectors within each superposition component with a single vector, and increase the number of dimensions of the vector space accordingly. The 3ND analogue of $\Psi_8$ looks like this:

$$\Psi_8^* = \sqrt{1 - \alpha^2} [(6_1|0_2|9_4|0_5|7_6|...|0_{3N}) > ] + \alpha [(0_1|2_3|0_4|5_5|0_6|...|1_{3N}) > ]$$

As we saw, one cannot deduce any macroscopic objects simply by considering this description. One could of course assume that every third dimension corresponds to the x-dimension etc., and then deduce macroscopic objects accordingly. But there is no privileged way of doing this: the quantum state does not motivate this. So the only way that the phenomenal analysis can help us move from such states to macroscopic truths about localised particles is if we postulate substantive psychophysical principles. An example of such a principle would be ‘if $\Psi_8^*$ obtains then there is an experience of a left result’. But the problem is that we have no independent reason to believe in such complicated psychophysical principles, and so we are left with no reason to believe this version of the collapse theory. Having said this, this is not a positive objection to the empirical adequacy or
coherence of the theory. And so this is the most plausible version of the collapse theory we have seen so far. And if one is sceptical of the phenomenal analysis of spatial concepts, one can simply abandon them, which leaves us with *Eliminativism*, which I now explain.

**Against Eliminativist Wavefunctionism**

Instead of explaining a particular phenomenon (e.g. consciousness, localised objects), eliminativism denies the existence of the phenomenon and sets out to explain why we wrongly thought the phenomenon to exist. This is usually done by explaining why we have non-veridical experiences of the phenomenon in terms of psychophysics.

Eliminativist collapse theories avoid the tails problem by filling all branches of the wave function (particularly the tails) with quantum mechanical *we know not what*. This can be done, as we previously saw, by moving to the 3ND view, for example by rewriting $\Psi_8$ as:

$$\Psi_{8\ast} = \sqrt{1-\alpha^2} \left[ |(6_1|0_2|0_3|9_4|0_5|\cdots|0_{3N}) > \right] + \alpha \left[ |(0_1|0_2|2_3|0_4|5_5|0_6|\cdots|1_{3N}) > \right]$$

Now the eliminativist just needs to postulate principles relating such quantum states to non-veridical experiences. The only difference between this view, and the view just described, is that eliminativism denies that spatial concepts can be analysed in phenomenal terms, or in any other terms that would enable a reduction of the relevant spatial truths to quantum truths. For this reason, eliminativism faces the same problem: we have no independent reason to believe in such complicated psychophysical principles, and so we are left with no reason to believe this version of the collapse theory. Again, this is not a decisive objection, and so either eliminativism or the phenomenal analysis view is the most plausible view yet. I now consider the claim that expanding the fundamental ontology provides us with something better.

**Against Wave Particle Dualism**

A number of philosophers (including some of the originators of the GRW theory) have thought that the best solution involves expanding the fundamental ontology. The idea is to add to the wavefunction ontology additional quasi-classical entities that are better suited to ground macroscopic truths. The new ontology is related to the wave function by new fundamental laws of nature. The simplest example of such a theory is the mass-density theory, which introduces a pattern of three dimensional
mass-density lawfully correlated with the wave function. The theory introduces mass density operators:

\[ m(r) = \sum_k m_k N^{(k)}(r) \]

Here \( N^{(k)}(r) \) is a particle number operator that gives the number of particles of type \( k \) that exist at position \( r \), and \( m_k \) is the mass of a particle of type \( k \). The mass density function at \( r \) is then defined as:

\[ M(r) = \langle \Psi | m(r) | \Psi \rangle \]

Here \(|\Psi\rangle\) is the quantum state vector of the universe. Flanking the density operator with two instances of the quantum state vector just means that we are weighting the mass density by the modulus square of the amplitude assigned to that region. As Wallace (2008: 60) puts it, “the mass density is the sum of the mass-weighted ‘probability’ densities for finding each particle at \( r \); for a one-particle wavefunction \( \Psi(r) \), \( M(r) \) is just \( m \times |\Psi(r)|^2 \).

The new fundamental law therefore states that the very existence of the wave function \( \Psi \) determines the existence of the mass distribution \( M(r) \). If this fundamental law is acceptable, then formulating a [Composition] expansion conditional should be no more problematic than in the classical case: we just group bits of mass density in thought and label those bits 'composite such and such'. If the mass density is correlated with a suitable high amplitude branch, then we will be able to derive everything we wish to explain, or so it seems.

Before evaluating whether this theory solves the tails problem, note that the mass-density principle is a departure from ordinary scientific theorizing, as noted by Monton:

“It’s actually the case that the dynamics of spontaneous collapse theories only depends on the quantum state; the evolution of the quantum state does not depend on the mass density. Instead, mass density is epiphenomenal: the mass density at a particular time is determined by the quantum state at that time, but the mass density does not have any influence on the future evolution of the system.”

(Monton 2004: 419)

Epiphenomenalism is familiar from the philosophy of mind, where some consider the physical universe to be causally closed, but consider consciousness to be a further fundamental entity whose

---

81 This view was introduced by Ghirardi, Grassi and Benatti (1995). The following simplified description of the view comes from Monton (2004) who distinguishes mass density simpliciter (presented here) from the problematic notion of accessible mass density, discussed in a moment.
states are determined by the physical state.  

Epiphenomenalism about consciousness is often defended as the last resort: we cannot explain consciousness in terms of the causally closed physical world, but we also cannot eliminate it because we have subjective introspective knowledge of it.

The mass-density view appears worse off, however, as it does not seem to be a last resort due to the availability of Bohmian mechanics and the Everett interpretation (see next section). Furthermore, there are also the two views from above that postulate substantive psychophysical principles (eliminativism and the phenomenal analysis view). Note that neither of these two views even makes sense in the philosophy of mind case: we cannot eliminate consciousness because we have direct introspective knowledge of it, and we cannot analyse consciousness in terms of what causes consciousness. So it is not clear why one should believe the mass density view.

Maudlin (2007) defends wave particle dualism, and particularly the mass density view. Interestingly, after dismissing a posteriori wavefunctionism, Maudlin spends much time arguing for why dualism is superior to a view that postulates psychophysical principles. Following Bell (1987), Maudlin argues that physical theories require local beables. Local beables are physical objects entailed or predicted by a fundamental physical theory, which play a crucial role in the testing of the theory. In particular, they let physical theory avoid postulating psychophysical laws:

“[T]here is no doubt a question about how, when we look at a rock with a certain shape, a conscious experience of a certain sort arises, but classical physics could put that question off for another day (which perhaps would never come). All classical physics needs is the belief that experiences as of a rock of a certain shape typically are experiences of a rock with that shape, and the physics could take care of the rock. It is hard to see even how to begin if the physics has, in its own terms, to take care of the experiences”.

(Maudlin 2007: 3260-1)

Maudlin continues:

“In sum, a physics devoid of local beables would be a radically different kind of physics, a physics faced with problems of a completely different scale and sort than any theory in human history. [...] It would be a change in the physical account of the world infinitely more staggering than, say, the addition of a few compactified dimensions to spacetime, or the

---


83 While a kind of epiphenomenalism is happening in Bohm’s theory, it is not as drastic. The evolution of Bohm’s additional variables—the corpuscles—is determined not just by the quantum state, but also by the initial positions of the corpuscles, at least on a common way of understanding Bohm’s theory.

84 Though this hasn’t stopped some philosophers e.g. Rey (1983) and Dennett (1988).

85 Maudlin (2007: 3161-2) cannot see how the relevant “local beables” could be an analytic consequence of the wave function. It is worth noting that Maudlin thinks the wavefunction is 3N-dimensional, which provides further motivation for wave particle dualism.
admission of a physical foliation of spacetime, or the understanding of electrons as states of small vibrating strings, or the introduction of a discrete spacetime structure at the Planck scale”.

(Maudlin 2007: 3161)

With these arguments in place, Maudlin rejects dynamical collapse theories that do not expand the ontology beyond the wave function, and goes on to develop the mass-density view, as well as a related view.

In response, there is nothing wrong with postulating psychophysical principles, at least if they are motivated to some extent. In general, if an otherwise adequate fundamental theory fails to explain phenomenon p, then we should consider the possibility that our experiences as of p are illusionary, and that there are psychophysical laws that explain such illusions. If there is no reason to accept the psychophysical laws, then we should simply abandon the theory. This is standard practice in science.

Consider the example of solidity. Our immediate experiences suggest that ordinary objects are everywhere dense: everyday observation finds no holes in rocks and so forth. But then molecular physics comes along, which is an otherwise adequate theory, but which renders these experiences non-veridical. For according to molecular physics, ordinary objects are mostly empty space and are utterly full of holes. What is the correct response to molecular physics? Here is one possible response: we can retain the postulates of molecular physics but avoid psychophysics by postulating *continuous beables* that are nomically related to the objects of molecular physics through epiphenomenal laws. In particular, we can retain the idea that experiences *as of* a continuous lump of matter is an experience *of* a continuous lump of matter, which is predicted by our theory of molecular physics supplemented by continuous beables. Clearly this is not our actual reaction to molecular physics. Our actual reaction is to develop psychophysics, perhaps by developing our theory of how light is reflected by the disperse molecules that compose ordinary objects and how our eyes and brains interpret that light. The idea that we should instead postulate continuous beables goes against ordinary scientific practice. But what exactly is the difference between postulating continuous beables and local beables? Why is the former bad but the latter good?

The examples from the history of science quickly pile up. Rather than eliminating colours (as direct realists conceive them) and developing the psychophysics of colour perception, should we instead postulate colour beables? Rather than eliminating the flow of time in response to (the block-universe interpretation of) relativity theory, should we postulate flowing beables? In all such cases, postulating psychophysical principles looks to be the more plausible option.

---

Newton (1672) himself preferred colour eliminativism: “to determine more absolutely, what Light is, after what manner refracted, and by what modes or actions it produceth in our minds the Phantasms of Colours, is not so easie.”.
There are much greater problems with wave function dualism. It still faces the tails problem, as recognised by Cordero:

“Thus, the above mass density formulation cannot help in a Schrödinger’s cat situation, for two reasons: (a) As the above expressions [definitions of \( M(r) \)] make plain, the mass density function respects the internal architecture or ordinary macroscopic objects; and (b) the center of mass density function [...] remains “tailed”’.

(Cordero 1999: 67)

I take the idea of the mass density function respecting the internal structure of macroscopic objects to mean that although the objects in the tails are “light” they are nonetheless structured just as the “heavy” objects determined by the high amplitude superposition component. This is essentially the symmetry argument: if advocates of mass density commit to the existence of the “heavy” objects in virtue of the fact that they are defined by the mass density function, then by symmetry they should be committed to the “light” objects. But the “light” objects are structured like measuring instruments and people. The “one-world” constraint on dynamical collapse theories is still violated, and the theory is a priori inconsistent.

Maudlin only briefly considers the tails problem in his defence of mass density:

“These difficulties arise if one tries to hold to the rule that physical facts exactly correspond to facts about which operators the wavefunction is an eigenstate of, and with which eigenvalues. The so-called ‘tails’ problem undercuts the utility of this rule”.

(Maudlin 2007: 3156 n3)

The “eigenstate-eigenvalue link” that Maudlin refers to is just the strict version of the fuzzy link principle (where \( p=0 \)). What Maudlin does not here recognise is that the mass density link makes the tails problem even worse: where the fuzzy link principle tries to (unsuccessfully) define away the tails, the mass density link defines the tails into the ontology87 (with the caveat that they are not very “massive”88).

In their original paper, Ghirardi, Grassi and Benatti (1995) in fact seem aware of this problem and try to define away the mass density tails with a further principle. They simply define light weight tails as “inaccessible” and “not objective”. Monton (2004) criticises these obscure pronouncements and

87 In a later paper, Maudlin (2010: 138) appears to agree: “at the end of the day, we need in addition an association of the world we experience with the high-density world, and the consequent neglect of the low-density worlds, and I don’t see how that is to be accomplished”.

88 If the analysis of ‘mass’ in section 2.2 is correct then this primitive ontology can’t literally be a mass density. This point is recognised by Allori et. al. (manuscript: 3.2): “the matter that we postulate and whose density is given by the \( m \) function does not ipso facto have any such properties as mass or charge; it can only assume various levels of density. For example, the \( m \) function is not a source of an electromagnetic field [...] we postulate that the primitive ontology consists of only one density—the matter density.”
defends the mass density simpliciter view (against the “accessible mass density” view). But to get around the light weight tails, he postulates psychophysical principles:

“I admit that, for the mass density simpliciter link to solve the tails problem, a certain assumption about psychophysical parallelism needs to be made. But the assumption is a reasonable one. [...] There is no reason to suppose that mental states supervene just on particle location; instead we can suppose that mental states supervene on the distribution of mass. Since the masses of particles in a brain are concentrated in the appropriate regions of space, it is reasonable to assume that the appropriate mental states supervene on those mass concentrations.”

(Monton 2004: 418)

To understand this proposal, reconsider the post collapse state (given in 3D terms):

$$\Psi_B = \sqrt{1 - \alpha^2} \left[ |(\text{PERCEIVE LEFT}) >_E |(\text{LEFT}) >_M |(6_x 0_y 0_z) >_P \right]$$

$$+ \alpha \left[ |(\text{PERCEIVE RIGHT}) >_E |(\text{RIGHT}) >_M |(0_x 0_y 2_z) >_P \right]$$

The idea is that mental states supervene, or are grounded, not only in particle configurations, but in the extent of their mass density. And so because the configuration that composes the experimenter perceiving a right result is “heavy”, there is only an experience of a right result, and not an experience of a left result. But all that does is leave us with zombie tails: tails containing objects that (other than being “light”) are physically identical to humans but are unconscious. This is not an acceptable theory.

I have used AET to evaluate the various attempts to make dynamical collapse theories work. Assuming that such theories are by definition “one world” theories, I have shown that almost all of them are internally inconsistent. The only versions that were not shown to be internally inconsistent were versions that postulate substantive psychophysical principles. But because these principles cannot be independently motivated, there is no reason to believe these versions of the theory. I thus conclude that dynamical collapse theories are false. In the final section, I discuss the significance of this result for AET and for quantum metaphysics more generally.

4.4 Quantum Metaphysics

Monton also tells us (2004: 415-6) on the basis of personal communication, that Ghirardi prefers mass density simpliciter.
I have argued that AET is supported by Newtonian and relativistic physics. I then assumed the truth of AET and used it to undermine a particular range of quantum theories—dynamical collapse theories. A critic might respond that we should take the inconsistency of AET and dynamical collapse theories to instead show that AET just a false relic of a pre-quantum world view. I think there are a number of reasons for why this is not the correct reaction.

Firstly, I have already given reasons to doubt this approach. In sections 2.9 and 3.4 I argued that the classical case studies support a very general metaphysical picture in which fundamentality is a property of descriptions and grounding is an asymmetric relation among descriptions. If this view is right then it is plausible that proposition P is grounded in Q if and only if P is a coarse grained description of the element of reality that Q is a fine-grained description of. The asymmetric nature comes from the fact that the coarse grained description can be inferred from the fine grained description but not vice versa. It is hard to see how quantum mechanics could undermine such a picture. The picture is instead motivated by its ontological simplicity and overall theoretical unity. It’s simple because it postulates only a single level of reality rather than multiple levels and it is theoretically satisfying because it enables natural analyses of the metaphysical notions of grounding and fundamentality. It is hard to see why quantum mechanics should be taken to challenge this plausible thesis.

Secondly, I argued in section 3.3 that higher level reductive explanations support AET. It is plausible that successful high level reductions, such as the reduction of solidity in terms of chemistry, are not going to be undermined by quantum mechanics. For surely we want quantum mechanics to explain (i) the existence of molecules and (ii) the nature of their solidity-generating relations. If a quantum theory cannot do so, then arguably it is not adequate. So if such high level reductions support AET, and AET is inconsistent with a class of quantum theories, then we should be sceptical of that class of quantum theories.

Thirdly and perhaps most importantly, insofar as section 4.3 was an illuminating critique, one could take the result of 4.3 as an argument in favour of AET in itself. I think the critique of section 4.3 is an illuminating one in an important sense. Some might worry that the method of critique was too philosophical, too aprioristic, and not sufficiently naturalistic. Surely, if anything is going to rule out dynamical collapse theories it would be an experiment and not “armchair theorising”. But this misunderstands the form of the critique: the application of AET to dynamical collapse theories has illuminated why such theories are empirically false: dynamical collapse theories are empirically refuted by the fact that physical systems do not spontaneously heat up.

Recall that by the uncertainly principle, the more concentrated a wave function’s amplitude for position is, the more dispersed its amplitude for momentum. But the more dispersed momentum amplitude is the more energy the system can possess. So the question of the empirical adequacy of
dynamical collapse theories is the question of whether one can prevent severe violations of energy conservation in such a way that one avoids an Everettian many worlds ontology. The AET analysis of the tails problem shows that this is probably impossible, so we should reject dynamical collapse theories on these empirical grounds. Advocates of dynamical collapse theories are therefore guilty of postulating a many worlds ontology, while insisting that one shall not call it a many worlds ontology. AET is useful here because its hard-line emphasis on conceptual analysis enables us to better identify such fallacies.

Since the conceptual difficulties in quantum mechanics runs so deep, such fallacies or easy to fall into, and theorists are not to be blamed to any great extent for making them. This just shows how essential AET is to quantum metaphysics. I will therefore conclude by suggesting areas of quantum metaphysics in which AET should be applied, and where I, and I hope others, will be applying AET in future research.

The two most popular interpretations of quantum mechanics other than dynamical collapse theories are many worlds theories, first formulated by Everett (1957) and additional variables theories such as Bohm’s (1952). Both types of quantum theory face a number of grounding problems.

The many worlds theory faces two primary problems. The first is whether the worlds can be adequately defined, or better, whether they are actually a priori deducible from the wave function. In the context of evaluating dynamical collapse theories I argued that the worlds (the tails) are deducible. But that’s partly because dynamical collapse theories ontologically privilege a basis (e.g. the position basis) in which the collapse occurs. The many worlds interpretation does not privilege a basis in this way: the position basis or the momentum basis etc. is meant to be equally good descriptions of reality. But such descriptions appear to disagree on the worlds. This problem is usually solved by appeal to decoherence theory. But decoherence alone cannot solve the problem as decoherence does not quite select a definite basis. This is where frameworks like AET can help. Indeed, the extra theoretical apparatus that contemporary Everettian’s appeal to, appear to be theories of grounding that are much less precise than AET.

The second problem faced by many worlds theory is the probability problem. If every possible measurement outcomes occurs (in some world) after a measurement, then it looks as though all measurement outcomes should be assigned probability one. But quantum mechanics does not assign all outcomes probability one, therefore many worlds theory is not an adequate interpretation of quantum mechanics. The solution to the problem often takes the form of a decision-theoretic proof.

---

90 See Schlosshauer (2008: section 8.2.1) and Butterfield (2001: section 7).
92 The problem is forcefully stated by Albert (2010).
that relies on controversial analyses of probabilistic concepts such as chance and credence. Evaluating this proof using AET would require that one determines whether the axioms of the proof are a priori or follow a priori from the quantum state.

Finally, Bohmian mechanics raises a number of issues. One of the problems is much like the tails problem and is sometimes called the “Everett-in-denial” problem. Bohm does not eliminate the wave function but adds to it (hence “additional variables”). Bohm introduces new entities associated with one branch of the wave function and asserts that these new entities compose manifest reality (much like the mass density view). The theory is supposed to be a “one world” theory and so a problem arises if one can deduce many worlds from the wave function. Whether this is the case will depend on careful analysis of how the wave function is formulated in Bohmian mechanics.

---

93 See Wallace (2010a).
94 The objection to Bohmian mechanics is put forward by Brown and Wallace (2005). For discussion see Lewis (2007), Brown (2010), and Valentini (2010).
References


Dennett, Daniel. 1988 “Quining Qualia”. In: Consciousness in Contemporary Science, edited by A. Marcel and E. Bisiach. New York, Oxford University Press.


Witmer, Gene. 2008. Review of Steven Horst’s “Beyond Reduction: Philosophy of Mind and Post-Reductionist Philosophy of Science”. In Notre Dame Philosophical Reviews [Electronic Journal].