UNIVERSITY OF CALIFORNIA,
IRVINE

A Categorical Consideration of Physical Formalisms

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Philosophy

by

Sarita Dalya Rosenstock

Dissertation Committee:
Professor James O. Weatherall, Chair
Professor Jeffrey A. Barrett
Professor JB Manchak

2019

DEDICATION

To Jim,
for believing in me so thoroughly
that my self-doubt could not compete

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGMENTS

Chapter 2 and 3 are substantially excerpts and reprints of Rosenstock et al. (2015) and Rosenstock and Weatherall (2016), respectively, with permission from the copyright holders. I am extremely grateful for my coauthors on these works, James Owen Weatherall and Thomas William Barrett, for a fantastic collaborative experience.

During my graduate studies at UC Irvine, I have received support from a number of sources. I have received merit fellowships and teaching assistantships from the School of Social Sciences, as well as a fellowship in honor of Christian Werner. I have received support from the National Science Foundation from grants # 1331126 and # 1328172 with James Weatherall as PI, and grant # 1535139 with Cailin O'Connor. For the 2015-2017 academic terms, I was supported by the Philosophy of Science Association for serving as managing editor of the journal *Philosophy of Science*. I am grateful to all of these funding sources, and particularly to James Weatherall for his aid in securing them.

This work would not have been remotely possible without the incredible support of my committee, composed of Professors James Weatherall, Jeff Barrett, and JB Manchak. Together they provided me with an incredible, incomparable philosophy of physics training. I cannot thank Jim enough for the amount of support he has provided me from the very beginning of my studies here, as an advisor and teacher and collaborator and friend. I am grateful to the many other academics who have helped me work through ideas over the years, in the UCI philosophy of physics reading group, as well as the many conferences where I presented parts of this work. And to Michelle Feng, who piqued my interest in topological data analysis and tolerated my prodding questions.

I would also like to thank my friends and family for their love and support these past 7 years. Special thanks to Will Stafford, for helping me stay focused and motivated in the final stretch, as well as formatting my bibliography when my wrist was out of commission. A final thanks to Isaiah, my partner in all things, who has upended his life twice over to stay by my side as I continue my academic journey.

CURRICULUM VITAE

Sarita Dalya Rosenstock

APPOINTMENTS

Research Fellow, School of Philosophy — **from 2019**

Australian National University — *Canberra, Australia*

EDUCATION

PhD in Philosophy, LPS — **2019**

University of California, Irvine — *Irvine, CA*

Master of the Arts in Social Science, IMBS — **2016**

University of California, Irvine — *Irvine, CA*

Bachelor of Arts in Mathematics and Philosophy — **2012**

Bard College at Simon's Rock — *Great Barrington, MA*

REFEREED JOURNAL PUBLICATIONS

Why Be Regular? Part I (with B. Feintzeig, J. Manchak, & J. Weatherall) — **2019**

Studies in the History and Philosophy of Modern Physics

When it's Good to Feel Bad: Evolutionary Models of Guilt and Apology (with C. O'Connor) — **2018**

Frontiers in Robotics and AI

In Epistemic Networks, is Less Really More? (with J. Bruner & C. O'Connor) — **2017**

Philosophy of Science

A Categorical Equivalence between Generalized Holonomy Maps on a Connected Manifold and Principal Connections on Bundles over that Manifold (with J. Weatherall) **2016**

Journal of Mathematical Physics

On Einstein Algebras and Relativistic Spacetimes (with T. Barrett & J. Weatherall) **2016**

Studies in the History and Philosophy of Modern Physics

Annealing of Directionally Solidified Alloys Revisited: No loss of solidification texture in Earth's inner core (with Y. Al-Khatatbeh, M. Bergman, D. Lewis, Z. Mason, and L. Zhu) **2016**

Studies in the History and Philosophy of Modern Physics

ABSTRACT OF THE DISSERTATION

A Categorical Consideration of Physical Formalisms

By

Sarita Dalya Rosenstock

Doctor of Philosophy in Philosophy

University of California, Irvine, 2019

Professor James O. Weatherall, Chair

In the progression from Newtonian physics to general relativity, the structural feature of absolute rest was abandoned because it was not necessary to account for the empirical validity of Newtonian physics. Ockhams razor-type arguments like this one, which appeal to a desire for minimal ontologies and more unified physical laws are often invoked in favor of one theory or model over another. But how do we distinguish between essential "structure" in a theory, and inessential contingencies of a particular description? Is there a precise way to "compare structure" across theories expressed with different language and mathematical constructs?

In my dissertation, "Structure and Equivalence in Physical Theories," I adopt and adapt a method of comparing the structure of different formalisms that I call "theories as categories of models" (TCM), recently suggested by J. Weatherall (2016) and H. Halvorson (2012). The motivating idea is that information about relationships between formal theoretical models provide crucial insight into the way in which these models are intended to represent real-world systems. Incorporating this so-called "functorial" information into the presentation of a theory naturally yields a mathematical object called a category. Formal methods from category theory can then be used to enrich our understanding of the nature of these models and the systems they represent.

While I aim to develop a rich and rigorous account of TCM, I focus primarily on the ways in which it has been and can be productively employed by scientists and philosophers, rather than merely considering this method in the abstract. I couch my analysis in three primary case studies in which TCM has produced novel insights into physical theories. Two of these applications, in general relativity (2015) and Yang-Mills theory (2016), are based on original theorems that I proved and published with co-authors, establishing that formalisms that many theorists consider meaningfully distinct are in fact equivalent in a precise, category theoretic sense. In the dissertation, I present these examples with a closer eye towards explicating the role that TCM plays in scaffolding the arguments in these papers. In the second case, I demonstrate how TCM opens the door to a larger and richer space of possible Yang-Mills formalisms, and indicate how the category theoretic structure in this space reveals the relationships between quantum field theories based in different classical formalisms.

I also consider topological data science (TDA), a popular method of analyzing the "shape" of large data sets. This represents a departure from other philosophical work on TCM, which has focused on theoretical foundations of theories from physics, whereas TDA is employed by a variety of researchers in multiple fields at a less theoretical, more practical level. TDA is a promising candidate for TCM because category theory is invoked by data scientists themselves to justify the use of its core methods. This case study reveals how scientists are and can be motivated by category theoretic considerations, and the ways in which these motivations do and do not align with those of philosophers of science employing the same tools. This chapter points towards new ways philosophers of physics might enhance TCM by analogy with TDA. In the other direction, the philosophical framework of TCM enhances the story data scientists want to tell about how TDA gets at the underlying "structure" of data.

Chapter 0

Introduction

This dissertation involves a variety of disciplinary aims and methods. The original motivation for the project comes from the philosophy of physics. I was unsatisfied with the reasons both physicists and philosophers gave for why one formalism should be preferred to another as a foundation for Yang-Mills theory. In this I am aligned with the burgeoning sub-field of philosophy of modeling, which seeks to clarify the diverse and puzzling roles that mathematical models play in mediating scientific understanding of the world. I was hung up, however, on the more fundamental philosophical issue of how to identify and individuate the formal models themselves, independent of arbitrary linguistic and stylistic choices made in their presentation. This quest to extract meaning from symbols aligns me with philosophers of language. My background in mathematics biases me towards a formal answer to this question, inspired by the work of logicians to connect syntax to semantics. Unfortunately, the messiness of how physicists employ formal models did not properly align with the sterile precision of logic. To accommodate scientific practice, models need to be imprecise enough to encompass a variety of possibilities, but capable of being sharpened when careful distinctions are warranted and useful. Scientific models are dynamic—evolving and adapting to changing goals and evidence. I needed a way to sketch the outlines of physical formalisms without holding them too tightly, but still being clear and consistent.

These considerations led me to *category theory*, a field of mathematics designed to characterize abstract mathematical structures. Category theory is often presented as an alternative to set theory which describes mathematical concepts abstractly by the role they play in mathematics rather than by concretely building them out of sets.[1] In doing so, it allows one to work with a mathematical concept without conceiving of a particular instantiation of it—the abstraction underlying equivalent descriptions. Category theory reveals core similarities between seemingly different ways of describing mathematical formalisms. In this text, it serves a scaffold for presenting formalisms, and conducting analyses and debates about whether and how one may be preferable to another for a given representational purpose.

This is the backdrop of the first project I engaged in associated in this dissertation, presented in chapter 3. The goal was to present "alternative" Yang-Mills formalisms as categories in order to bring clarity to current debates in the foundations of physics. In practice, this involved doing much more differential geometry (the branch of mathematics used to formulate Yang-Mills theory) than category theory. The category theory served to frame the questions and summarize the results, but the "hard part" involved employing the formalisms in the language they are presented (differential geometry) in order to fit into this simple category theoretic framework. Category theory was an essential part of the context of discovery, but was arguably eliminable from the ultimate presentation and justification of the results. This made its utility in the project difficult to pin down. The insight it yielded into the foundations of physics was interesting in its own right, but I was also curious as to *how* those insights were achieved using category theory. Understanding this required applying the technique—which I refer to here as "theories as categories of models", or TCM—to other

[1] This text does not take a stance on the debate in philosophy of mathematics as to whether category theory or set theory better serves as a foundation for mathematics. One can view mathematical objects as being concretely built from sets while still benefiting from a category theoretic perspective on them.

questions in the philosophy of science. The second project I pursued, presented in chapter 2, employs TCM to adjudicate a debate on the foundations of general relativity. This case study did provide additional insight into how TCM works, but it was still too similar to the Yang-Mills case to reveal the full scope of what TCM has to offer. This inspired a third project, discussed in chapter 4, which presents a novel perspective by examining the role of category theory in data science. It is with the knowledge acquired from these three applications that I developed the account of TCM you will read in chapter 1.

Chapter 1 is in a sense the culmination of this work, but also its starting point. Here I present both the motivation and ultimate justification for the "theories as categories of models" framework employed in the chapters that follow. This framing contrasts with the presentation of TCM as other philosophers of physics see it (Weatherall, 2017; Barrett and Halvorson, 2016; Hudetz, 2019, for example.). In these works, establishing an (in)equivalence of categories (that meets certain criteria) is presented as a method for establishing when two theories are (in)equivalent to one another. A fully adequate presentation of TCM would then involve offering necessary and sufficient conditions for an equivalence of categories to serve this role in philosophy of physics. Such conditions include "commuting with observations" (Weatherall, 2017) and "definability" (Hudetz, 2019). In contrast, I view TCM as a way of presenting a pre-established account of how physical formalisms relate to one another and the systems they represent. The only restrictions I take TCM to impose are what is required for such an account to be internally consistent. The role of TCM, I argue, is to reveal the commitments and entailments of the presumed relationship between formalisms, and expose them to further analysis.

I proceed in chapter 2 to apply TCM to demystify a proposed alternative formalism to general relativity. I discuss this application first because it is the easiest to both comprehend mathematically and motivate philosophically. This work responds to John Earman's (1986) proposal that Robert Geroch's (1972) "Einstein algebra" formalism might provide a way to present general relativity without the "spacetime substance" that he claims is present in the traditional formalism of manifolds with metric. A novel theorem is proven, demonstrating a category theoretic sense in which Einstein algebras should be understood as having the same structural content as manifolds with metric, undercutting Earman's argument. While on my view of TCM this does not definitively establish that Earman's approach cannot work as intended, it shifts the burden of proof to Earman to present a TCM narrative that does function as he intends, which there is reason to believe is not viable.

It is not until chapter 3 that I present the project that inspired this dissertation. Here I use TCM to illuminate the relationships between different ways of formalizing Yang-Mills theory, a generalization of electromagnetism that, in its quantized form, underlies much of modern particle physics. This application is much messier, but arguably more illuminating and relevant to practicing physicists than the previous one. This work responds to Richard Healey's (2007) argument that formulations of Yang-Mills in terms of what are called "holonomies" are preferable to formulations in terms of "principal bundles," largely on the basis of parsimony considerations. Healey claims that principal bundle formulations posit "surplus structure" relative to holonomy formulations (p. 30), so we should expect that the latter captures the true structure of the world, whereas the former possesses unnecessary mathematical fluff that obscures the physical interpretation. I prove a novel theorem demonstrating a category theoretic sense in which the holonomy formalism can be taken to in fact posit *more* structure than the principal bundle formalism. I discuss how Healey's account can nonetheless be modified to fit the TCM framework.

Chapter 4 investigates a field of research—topological data analysis (TDA)—in which category theoretic considerations are already taken seriously by scientists in their modeling practices. TDA aims to identify the essential "structure" of a data set as it "appears" in an abstract space of measurement outcomes. At the heart of TDA is the concept of *homology*, an abstract mathematical interpretation of "hole" structure. Homology exhibits the category theoretic property of *functoriality*, meaning it is defined not only on models but on structure-preserving functions between them. There are ubiquitous hints in the TDA literature that its practitioners consider the functoriality of homology to be central to its utility in application, but this maxim does not appear to be explored in much depth. This discussion of TDA aims to provide insight into how scientists are and can be motivated by category TCM-esque considerations, and the ways in which these motivations do and do not align with those of philosophers of physics employing the same tools. I argue that the utility of category theoretic methods to researchers in this context is rooted in the particular geometric nature of the mathematical models. The category theoretic framework helps to connect topological models, which have straightforward physical interpretations, with algebraic models, which are more abstract but easier to process computationally.

This dissertation can be read in two different ways. One might take TCM as a means to an end, a tool whose value lies in its ability to help resolve the sorts of philosophical questions that the proceeding chapters address. Alternatively, one might view these applications as valuable insofar as they serve to articulate TCM, which is of philosophical interest in itself. In truth, this text is a product of both perspectives. Each application is of independent value to the philosophies of particular sciences, and in aggregate they provide window into the nature of scientific representation more generally.

Chapter 1

Theories as Categories of Models

1.1 Introduction

In this chapter, I introduce a philosophical program I call "theories as categories of models" (TCM), which forms the backbone of this dissertation. Variations on TCM have recently been deployed by a number of philosophers of science to analyze the content of theoretical formalisms and determine inter-theoretic relations in physics.[1] The central premise of TCM, as indicated by the name, is to use the mathematical concept of "category" to encode the formal content of physical theories.

TCM is applied in this text and elsewhere in situations where there is a question or disagreement about the relationship between two different theoretical formalisms that are used to represent the same class of physical system, such as a set of equations versus a geometric figure. In particular, such questions often follow the schema:

[1] For example, Weatherall (2016a,c); Barrett (2018a). See Weatherall (2017) for an overview of applications.

> Are formalisms A and B merely presenting the "same" physical facts in different packaging, or does one have "more structure" than the other in some sense?

In section 1.4, I will examine how understanding A and B as categories can aid in settling these sorts of questions, as deployed in chapters 2 and 3. Before I can do that, I need to address the more fundamental issue of how and why one should understand theoretical formalisms as categories in the first place. After all, answering questions about the relative content requires settling on what we are referring to when we talk about formalisms A and B.

The use of category theory by philosophers of physics is motivated by the observation that understanding the language independent content of each formalism requires considering not only the various ways in which the formalism can be used to represent a physical system (a collection of formal models), but also the ways in which such models can stand in relation to one another. Incorporating these relations between formal models—which include isomorphisms between models, and system-subsystem relations—provides crucial insight into the way in which these models are intended to represent real-world systems. Including these relations the presentation of a formalism leads us to construct a mathematical object called a *category*. Sophisticated tools from category theory can then be used to enrich our understanding of the nature of these models and the systems they are intended to represent.

This broad strokes summary is relatively philosophically inoffensive. What is more contentious is the stronger claim that the content of theoretical formalisms is *fully* captured by presenting it as a category. This stronger claim may appear to be necessary if one wants to, as I do in this text, settle questions about the relative structural content of formalisms on the basis of examining the relationship between categories associated with them. As I will argue in this chapter, however, such a strong commitment to a formal physical theory *as being* a particular category is not required for TCM to be useful, nor is it even recommended. I argue that we can (and should!) associate many different categories, as well as non-categorical information, with a given formal theory. Nonetheless, if we are sufficiently clear about the ways in which we purport to use formal mathematical objects to represent physical states of affairs in our theory, we can uniquely define a category that abbreviates this association of representations to states of affairs. I call this the *structure category* associated with a theoretical formalism. We can then compare these categories to resolve questions of relative structure without having to *identify* the physical theory with the structure category.

I begin in section 1.2 by explaining what I mean by "structure" in this text, and how considerations about structure motivate the presentation theoretical formalisms in category theoretic terms. In section 1.3, I describe a procedure for achieving reflective equilibrium regarding the structural content of formalisms. This procedure begins by laying out ones prior commitments regarding a formalism, the systems it represents, and how it is understood to represent them, and outputs a "structure category" which serves as a minimal summary of these prior commitments. In section 1.4, I introduce the "property-structure-stuff" (PSS) heuristic for evaluating the relative content of a pair of structure categories. I tie up some loose ends in section 1.5.1, and then conclude in section 1.6 with a more nuanced summary of TCM and its role in philosophy of physics.

1.2 From Structure to Categories

1.2.1 Structure

There are a few things that philosophers and physicists appear to mean when they talk about *structure* in a theoretical formalism. The term often refers to something along the lines of

Structure$_{\text{alt}}$: The relevant *causal* or *explanatory* content of a description, abstracted away from noise and unimportant details.

I will come back to this notion of structure in chapter 4 in discussing an application of TDA in data science. For most applications in the foundations of physics, however, including the ones discussed in this text, the operative notion of structure under discussion is more aptly characterized by

Structure: The *content* of the description, abstracted away from the particular words, symbols, or language used to formulate it.

This is closely related to another characterization of structure:[2]

Structure*: That which is *invariant* under (left unchanged by) re-description, or shared by equivalent descriptions.

The common thread in all of these is that structure plays a mediating role between a description and the phenomenon it describes. Structure$_{\text{alt}}$ gets at the idea that a formal framework, such as electromagnetism, can facilitate understanding and explanation even of systems we are measuring imprecisely, or in which there are other (less relevantly) operative phenomena present. The second two notions of structure—the focus of the majority of this text—get at the idea that electromagnetism is the same theory whether its in English or Urdu, no matter which symbols or fonts I use to write it down, or in which order I present Maxwell's equations. But the boundaries between content and description are not always known or agreed upon, and are sometimes themselves the subject-matter of scientific inquiry.

[2]The connection between structure and structure* is discussed in detail and expressed formally in Barrett (2018b).

1.2.2 Example: spacetime structure

When Newton formulated his theory of gravitation, he intentionally included a notion of absolute rest, and insisted that it was an ineliminable feature of the formalism.[3] This means that for Newton, two models of spacetime that differ only by a constant shift in velocity should be taken to represent distinct physical possibilities. Newton's view was controversial even at the time—Leibniz famously criticized it, and even before Newton was born, Galileo presented the following thought example to refute it. For someone in the hull of a ship, there is no experiment that could distinguish between whether the ship is in constant motion or at rest without looking outside the ship. By analogy, if we are considering the universe instead of a ship, there should be no way to distinguish between rest and constant motion, as there is no "outside" perspective from which to make that distinction. Nonetheless, Newton's theory was the most comprehensive and empirically adequate available. And the debate over whether and how to eliminate absolute rest from spacetime theories persisted for centuries, and played a role in the development of physics during that time. It was not properly resolved until Einstein's theory of relativity provided a fully coherent and even more empirically adequate way to describe spacetime without absolute rest.

This disagreement appears to be rooted in the structure rather than presentation of spacetime theories. Critics of absolute rest are not (typically) troubled by the mere fact that the concept is *involved* in descriptions of systems—for instance, that objects are assigned velocities when describing a physical model. This is a notational strategy that physicists employ to this day, analogous to choosing a coordinate system to specify the location of objects in space. The contention lies with the further insistence that this feature is structural rather than merely notationally convenient. Such a commitment to absolute rest is captured by the claim that two spacetime models related to one another by a constant global shift in velocity, called a

[3] This account is simplified to serve the present philosophical aims, but should not be taken as proper historical scholarship.

Galilean transformation, represent distinct physical possibilities. Most physicists nowadays, in contrast, view Galilean transformations as merely changes in notational convention, and thus do not view such models as relevantly different from one another.

This example is a paradigmatic, clear cut case of two formal presentations of spacetime in which one posits more structure than the other. Barrett associates the two views with two different formal definitions of what spacetime is, and explains that *Newtonian spacetime* can be thought of as *Galilean spacetime* plus the extra structure of a preferred rest frame. Barrett (2015) argues that this relationship is captured by examining the *isomorphisms*—transformations of the formalism that preserve essential structure—that are admissible in each formalism.[4] All isomorphisms of Newtonian spacetime also preserve the essential structure of Galilean spacetime. But the reverse does not hold: there are many isomorphisms of Galilean spacetime that do not preserve the rest frame, and thus do not preserve the essential structure of Newtonian spacetime.

The idea here is that there is an intimate connection between the structure of a formalism and its isomorphisms. Barrett captures this connection more explicitly in (Barrett, 2018b), where he demonstrates that while isomorphisms are defined as structure preserving transformations, there is a sense in which isomorphisms can instead be taken to define what structure is. To add structure is to cease to consider as isomorphisms transformations that do not preserve that structure. To do so is to instead see such transformations as real changes from one possible state of affairs to another. To Galileo, a transformation that adds a constant velocity to all matter results in a state of affairs that is physically equivalent to the initial spacetime, and so constitutes an isomorphism. To Newton, in contrast, such a transformation results in a genuinely different physical situation.

[4]In this paper Barrett only considers *automorphisms*, or structure preserving maps from a thing to itself. I am extrapolating to his future work (2018b), in which Barrett establishes that automorphisms are not in general sufficient.

1.2.3 Categories

Barrett's argument that Newtonian spacetime has more structure than Galilean spacetime involves associating each formal theory with both a collection of formal models of spacetime, and a collection of isomorphisms between those formal models. In doing so, Barrett defines a category associated with each spacetime theory.

A *category* $\mathbf{C}$ consists of a collection of objects Obj($\mathbf{C}$), and a collection of morphisms Hom($\mathbf{C}$), or relations between the objects of the form $h : \mathbf{C}_1 \to \mathbf{C}_2$ with the following properties:

- For objects $A, B, C \in \mathbf{C}$ and morphisms $h_1 : A \to B$ and $h_2 : B \to C$, one can compose the morphisms as $h_2 \circ h_1 : A \to C$.
- Such compositions are *associative*, meaning $h_3 \circ (h_2 \circ h_1) = (h_3 \circ h_2) \circ h_1$
- Every object $C \in \mathbf{C}$ has an *identity morphism* $1_C : C \to C$.
- Identity morphisms leave objects unchanged in the sense that for any morphism $h : A \to B$, $h = 1_B \circ h = h \circ 1_A$.

When applied to theoretical formalisms, one considers categories in which the objects are formal models and the morphisms are formal transformations of those models. Most TCM applications in philosophy of physics, including chapters 2 and 3 of this text, look at only isomorphisms of the theoretical models. This yields a specific type of category called a *groupoid*.[5]

This use of category theory to understand the content of physical theories traces back to Halvorson (2012)[6]. This paper criticizes the *semantic view* of the nature of scientific

[5]In chapter 4 I discuss the benefits of including other relations in these categories as well, such as embeddings of subsystems into larger systems and projections to lower dimensional representations.

[6]See Halvorson and Tsementzis (2015) for a more recent presentation explicitly in terms of category theory.

theories, which identifies theories with classes of formal models. The semantic view purports to more closely capture the structural, language independent content of theories by enumerating models rather than focusing on how a theory is expressed in a particular language. Halvorson notes that merely enumerating models is insufficient to characterize the content of the theory, since relations between models are required to understand the full representation capacities of those models. There is a sense in which Newtonian and Galilean spacetimes involve the same *collection* of formal models of spacetime. It is only when one considers that the Newtonian picture assigns a special status to particular rest frames—which one can see by looking at the *relations between the models*—that the formal distinction is revealed.

When we talk about categories rather than collections of models, the relevant way of relating one category to another is a *functor* $F : \mathbf{C}_1 \to \mathbf{C}_2$ rather than a function. Functors are defined for both objects (models) and morphisms (relations between models), and it behaves well with those relations. That is, it takes identity morphisms to identity morphisms, and preserves morphism composition as $F(f \circ g) = F(f) \circ F(g)$.

Mathematicians have long used category theory to express abstract concepts in mathematics. Understanding physical theories as categories of models, and relations between theories as functors, enables philosophers and physicists to similarly apply the rich tools of category theory to understanding the abstract, language independent concepts underlying physical theories.

To explain why Galilean spacetime has more structure than Newtonian spacetime, we can define two categories—**Gal** and **Newt**—associated with each. In **Gal**, the objects are spacetimes[7] and the morphisms are Galilean isomorphisms. In **Newt**, the objects are those same spacetimes paired with an additional specified "rest frame", and the morphisms are the subcollection of Galilean isomorphisms that also preserve those rest frames. We then define a functor $F : \mathbf{Newt} \to \mathbf{Gal}$ which takes Newtonian spacetimes to their underlying Galilean

[7] See Barrett (2015) for precise definitions.

spacetime, thus "forgetting" the structure of of the rest-frame in a manner discussed more formally in section 1.4.

1.3 Category selection

The TCM framework can be briefly summarized by the claim that theories are categories of models, and relations between theories are functors between these categories. Of course, a lot of work is being done by the word "are" here. I doubt any of the philosophers employing TCM take themselves to be committed to fully identifying theories with a particular category. Rather, considering the way in which a particular formalism is used to represent physical systems tends to *suggest* a particular category, and the categorical presentation usefully summarizes an otherwise cumbersome collection of features of the formalism. But if one is not primed to be suggestible in this way, then the resulting association of theory to category—and subsequent inference from category to (relative) structural content—will be less than compelling. One might insist—and rightfully so—that there is more to the content of formalisms than can be expressed by a category of models.

Consider the spacetime example discussed above. To explain how Newtonian spacetime has more structure than Galilean spacetime, we define a functor $F : \mathbf{Newt} \to \mathbf{Gal}$ to capture the relationship between them, and point out that F "forgets the structure" of rest frames. This story is compelling largely because the claim that adding a rest frame adds structure to Galilean spacetime is relatively uncontroversial. But if you were not already convinced of this, you might dispute the choices I made for defining each of the categories associated to the formalisms, or the functor I chose to connect them, in order to evade the conclusion. Worse still, the same formalism might be reasonably associated with different categories. It is perfectly valid, and even sometimes recommended, that physicists working in a Galilean framework specify a rest frame and differentiate between isomorphisms that preserve it and those that do not. One does this in order to express the dynamical evolution of a point particle. The transformation from [spacetime + object at point x] to [spacetime + object at point y] is an isomorphism of spacetime, but even in the Galilean perspective, dynamical evolution of this form expresses genuine changes in states of affairs. Why should we privilege empty spacetime as *the* spacetime associated to a formalism?

Often, the informal procedure we have already invoked to define a category for a formalism (and the slightly more formal version described in 1.4.3) will be sufficient. However, in the most interesting and useful applications of TCM, like those in chapters 2 and 3, there is likely some disagreement or uncertainty regarding the relationship of relative structure of formalisms. In such cases, one might need to step back further and settle on how to define the formalisms they are working with, before they can ask how different formalisms relate to one another. In this section, I describe a procedure for achieving reflective equilibrium regarding which category to associate with a given formalism. This involves using category theory to lay out ones commitments, whatever they are, about the models they are using, the system they want to represent, and how they think the models hook up to the system. Importantly, this procedure will not tell you where these commitments should come from, merely how to present them coherently in order to define what I call "structure category." This category will have a special status as a kind of minimal summary of prior representational commitments. As such, it serves to provide motivation for why the category was chosen, and also exposes the commitments involved to analysis and criticism.

What we are after is a procedure for constructing a category associated with a particular theoretical formalism that captures its essential structure. Our starting point is a formal theory presented in some language with some symbols, and a general idea about the ways in which we can use formal objects to represent physical states of affairs. The output of our procedure should reflect the assumptions about the representational capacities of the formalism, and it should not depend on the language we used to write it down.

The nice thing about category theory is that it was invented to enable mathematicians to work with abstractions rather than particulars. In this way, I can use it to describe the outlines of how an act of representation can be understood, while allowing that this abstract outline be instantiated in any number of ways on the ground. So lets consider some minimal conditions for some formal thing x to represent a state of affairs ω.

In order for x to represent a physical state of affairs ω, x must be implicitly or explicitly selected from some collection **Rep** of representational choices, and ω must similarly belong to some collection of alternate states of affairs **SoA**. Representation also requires some kind connection $\iota : \mathbf{Rep} \to \mathbf{SoA}$ that "interprets" x as indicating ω in some way. A more complicated fact about language is that we almost never know exactly what these collections of possibilities or interpretations are supposed to be. Considerations about structure in the previous section indicate that we may additionally want to include relations between representational possibilities in describing **Rep**, and thus view **Rep** as a category rather than a collection.

At the extreme end, one might want to define **Rep** as consisting of all possible things one could use to represent a state of affairs, and **SoA** as all possible ways the world could be. I am not convinced that such definitions would be meaningful, and even if they were, they would certainly be unwieldy. Someone communicating or receiving information about ω via x will use some combination of context cues, pragmatic considerations, and explicitly stated rules to dramatically reduce the collections of possible states of affairs and representations to much smaller spaces that consist of the *admissible* ways of altering x and ω. These tend to shift over time and between contexts, becoming more or less inclusive and fine-grained as required for communication. This procedure assumes that what counts as admissible is given by external considerations, and constrains how "admissibility" is understood only by requiring internal consistency.

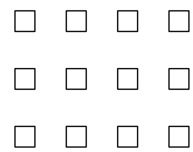


Figure 1.1: A configuration of chairs

1.3.1 Example: event planning

It helps to consider a concrete example of an act of representation outside of the more metaphysically loaded context of scientific theories. So lets assume $x =$ figure 1.1 is intended to represent a configuration of chairs at an event, a physical state of affairs indicated by ω. This representation is created by an event planner for their assistant to implement.

In theory, figure 1.1 has lots of features I might alter to generate a different representation in **Rep**. I could move it to the left, change its color, add more squares, add some triangles, etc. Similarly, there may be arbitrarily many ways of arranging the event room in the space **SoA**. But conventions about communication and event planning, and physical restrictions on how chairs work, will immediately reduce the scope of possibilities. It seems pretty clear, for example, that the event planner providing figure 1.1 intends to associate individual chairs in space with individual squares in space. So someone interpreting this diagram would reasonably read it as indicating that there should be 12 chairs, as opposed to 11 or 13, and moreover that these 12 chairs should be arranged evenly in 3 rows of 4 chairs. There is some room for reasonable people to debate, however, as to whether we should attribute significance to the spacing between the squares, and how neat the rows of chairs have to be for the arrangement ω to satisfactorily realize the schema indicated by x.

Neither the event planner nor their assistant will likely have in mind a formally defined representation $\iota : \mathbf{Rep} \to \mathbf{SoA}$. Their shared evolutionary past and social environments prime them to nonetheless implicitly act in accordance with largely compatible representations.

Miscommunication, when it occurs, is rooted in some discrepancy between the implicitly defined "interpretation functors" employed by planner and assistant. For example, the assistant might fail to pick up on the planner's intention to associate the color of the squares in figure!1.1 with chairs of the same color in the real event space.

While communication requires *some* way of associating representations with states of affairs, it is often the case that it is understood by both parties that there are alternative viable choices of interpretation on the table. For example, there is no indication in figure 1.1 as to which direction chairs should be oriented—though contextual considerations would reasonably limit the possibilities to ones in which the chairs are co-oriented and facing one of the four walls in the event room. So both planner and assistant might reasonably interpret figure 1.1 to be associated with multiple different representations $\iota : \mathbf{Rep} \to \mathbf{SoA}$. This might lead the assistant to ask for clarification—that is, ask the planner to select a particular ι—or interpret a lack of specificity as indifference on the planner's part as to which interpretation is implemented.

1.3.2 Representation diagrams

To summarize: in order for signals to contain information, signals and signified must be drawn from spaces of alternative possibilities, and there has to be at least one viable interpretation $\iota : \mathbf{Rep} \to \mathbf{SoA}$ from possible representations to possible states of affairs. Let us refer to the collection of admissible interpretations ι as I. I define a *theoretical representation diagram* as a choice of representation category **Rep**, states of affairs category **SoA**, and admissible interpretations I, as indicated by the diagram $D = \mathbf{Rep} \overset{\vdots}{\underset{I}{\to}} \mathbf{SoA}$. D presents a narrative about how a representation can be interpreted as physical states of affairs in a theoretical context, but the narrative itself is provided by external considerations about scientific theory and modeling practices. A theory may be associated with many different such diagrams at different times and in different contexts.

I do not claim that all acts of representation can be formalized as representation diagrams. I actually think most acts of representation cannot be straightforwardly interpreted in this way, but largely because most acts of representation are vague and ill-specified. Considering the various ways that an event planner might use figure 1.1 to indicate a spatial arrangement, it is unsurprising but impressive that successful communication can often nonetheless occur. But it can also break down, and these breakdowns can be explained by failures of precision in the act of representation.

Similarly, when a physicist refers to a particular mathematical object, the need for precision in specifying that object is less salient than it is for the mathematician. For example, when a mathematician invokes the "continuum" or the "real numbers," it is incumbent upon them to specify exactly which properties of the real numbers are required for their results to hold. At a minimum, this will typically involve the bare set of real numbers $\mathbb{R} = \{0, 329, \pi, \sqrt{2}, ...\}$. It may also involve an ordering relation $\leq$ on this set, which tells you for any two numbers which is bigger, or multiplication $\times$ or addition $+$ operations, or topological or distance structure. It generally matters to mathematicians which features of the real numbers are at work in a proof, and in what ways facts about the real numbers can and cannot be generalized to other mathematical objects. For physicists, in contrast, it is often sufficient that a mathematical construct serves its intended purpose. It is only when problems arise, or one asks theoretical questions about possible extensions and variations on a theory, that precise specification of mathematical constructs appears necessary.

Constructing a representation diagram is a useful way to achieve full precision, transparency, and consistency in how one presents an act of representation. Often, science does not require such precision. But TCM is intended for precisely those moments in theoretical development in which vague understanding of the referent of a formal apparatus proves insufficient.

1.3.3 The structure category

Let $D = \mathbf{Rep} \overset{\vdots}{\underset{I}{\to}} \mathbf{SoA}$ be a theoretical representation diagram. The *structure category of* D, written S_D, can be defined equivalently as either the coequalizer object $\mathbf{coeq}(D)$ of D or the quotient category $\mathbf{SoA}/\sim_I$ under an equivalence relation induced by I. I give the definitions of these concepts below, along with narrative explications of the phenomena they are supposed to capture. I do this in the hopes that one or the other will be easier to comprehend, and also because both properties of $\mathbf{S}_D$ are important for the analysis that follows justifying the coronation of $\mathbf{S}_D$ as a structure category.

Definition 1.1 (Coequalizer)**.** The *coequalizer diagram* of the diagram $D = \mathbf{Rep} \overset{\vdots}{\underset{I}{\to}} \mathbf{SoA}$ is a diagram $\mathbf{SoA} \overset{p}{\to} \mathbf{coeq}(D)$ which is such that $p \circ \iota = p \circ \iota'$ for all $\iota, \iota' \in I$, and is *universal* for this property.

That is, for any $q : \mathbf{SoA} \to \mathbf{C}$ which satisfies the same property that $q \circ \iota = q \circ \iota'$ for all $\iota, \iota \in I$, there exists a unique $q' : \mathbf{coeq}(D) \to \mathbf{C}$ such that $q' \circ p = q$.

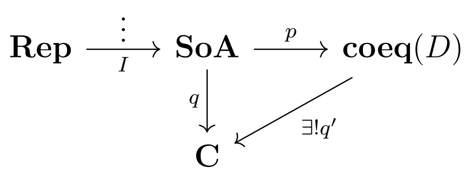


Figure 1.2: The commuting diagram of a coequalizer.

Coequalizers can be understood informally for present purposes as follows. A claim about a physical system can be thought of as an association $\omega \mapsto \sigma$ of some state of the world ω with some statement σ characterizing it. The argument in section 1.3 similarly applies here: for this association to be meaningful, σ must live in a category $\mathbf{C}$ of statements associated with other states of the world in $\mathbf{SoA}$, thus inducing a functor $q : \mathbf{SoA} \to \mathbf{C}$ which implicitly underlies any sufficiently well-defined connection between the state of affairs ω and the statement σ.

For such a claim $q(\omega)$ about a physical state of affairs ω to be rooted in my theory, it cannot depend on which of the admissible interpretations ι is used to connect the state of affairs ω to a theoretical representation in σ. That is, it must be the case that $q \circ \iota = q \circ \iota'$ for all $\iota, \iota' \in I$. Other statements might be true, interesting, relevant, etc., but they lie beyond the scope of the sense in which the objects in **Rep** represent states of affairs in the theory as specified by I.

Definition 1.2 (I-equivalent)**.** Given a diagram $D = \mathbf{Rep} \overset{\vdots}{\underset{I}{\to}} \mathbf{SoA}$, the I-equivalence relations $\sim_I$ on **SoA** is defined on objects as $\omega \sim_I \omega'$ if there is some object $x \in \mathbf{Rep}$ and functors $\iota, \iota' \in I$ such that $\omega = \iota(x)$ and $\omega' = \iota'(x)$. Similarly for morphisms $f, f' \in \mathbf{SoA}$, $f \sim_I g$ if there is some arrow $g \in \mathbf{Rep}$ and functors $\iota, \iota' \in I$ such that $f = \iota(g)$ and $f' = \iota'(g)$.

Definition 1.3 (Quotient)**.** Given a diagram $D = \mathbf{Rep} \overset{\vdots}{\underset{I}{\to}} \mathbf{SoA}$ and an equivalence relation $\sim$, the *quotient diagram* is the diagram $\mathbf{SoA} \xrightarrow{Q_\sim} \mathbf{SoA}/\sim$. The objects and arrows of $\mathbf{SoA}/\sim$ are equivalence classes of objects and arrows in **SoA**, and the functor $Q_\sim$ takes objects and arrows to their respective equivalence classes.

It follows directly from proposition 4.1 of Bednarczyk et al. (1999) that the I-quotient and coequalizer diagrams of D coincide. In this way we can think of $\mathbf{S}_D$ as states of affairs **SoA** "modded out" by equivalence with respect to the class of interpretations I of the formalism **Rep**. Importantly, this is *not* the same thing as quotienting by equivalence with respect to isomorphism of either states of affairs or representations. This allows us to preserve the idea that meaningful distinctions *can be* drawn between representations and states of affairs that are isomorphic to one another (and moreover that such distinctions can be an important component of the theory from which D is constructed), but decline to attribute those distinctions to the manner in which we are able to *interpret* representations *as* states of affairs, as indicated by the T-admissible interpretations I.

I define the *Ockham functor* associated with the theory diagram D as the functor $F_D : p \circ \iota$ for some interpretation $\iota \in I$. Note that since p is a coequalizer functor for D, this definition does not depend on the particular choice of ι. For a given $\iota \in I$, we can use the axiom of choice to construct a functor $\kappa_\iota : \mathbf{S}_D \to \mathbf{SoA}$ as $[x]_\sim \mapsto y \in [x]_\sim$ for any object or arrow $x \in \mathbf{S}_D$, where we require that $y \in \iota^{-1}(x)$ whenever this set is non-empty. Then for any interpretation $\iota \in I$, $\iota = \kappa_\iota \circ F_D$. I read this as an expression of the sense in which the interpretation $\mathbf{Rep} \xrightarrow{\iota} \mathbf{SoA}$, can be "filtered through" or "mediated by" $\mathbf{S}_D$ via the *Ockham diagram* $\mathbf{Rep} \xrightarrow{F_D} \mathbf{S}_D \xrightarrow{\kappa_\iota} \mathbf{SoA}$, as illustrated in figure 1.3.

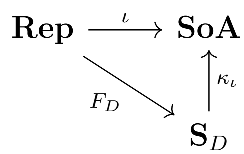


Figure 1.3: The Ockham diagram commutes with interpretations.

1.3.4 Spacetime revisited

At the end of section 1.2, I expressed the worry that deploying TCM seems to require arbitrarily fixing a particular category associated with a theoretical formalism. I now argue that this is not the case. Rather, I can grant that a given theoretical formalism can be accurately characterized by a variety of choices of the categories **Rep** and **SoA**, and interpretation functors ι. I instead claim that any particular use of a formalism implicitly fixes **Rep** and **SoA** and restricts the scope of "admissible interpretations" I, which can be summarized by the diagram $D = \mathbf{Rep} \overset{\vdots}{\underset{I}{\to}} \mathbf{SoA}$. It is my view that a shared structure category is a minimal requirement for identifying the same "formal structure" as it appears in various acts of representation. Moreover, I believe that the structure category captures what TCM practitioners have in mind when they associate a theoretical formalism with a particular category.

Suppose one objects that **Gal** is arbitrary, since **Gal***—generated by considering spacetimes with a single idealized particle, as discussed in section 1.3 above—could just as readily capture the content of the Galilean spacetime formalism. In this new framework, this amounts to the claim that **Gal*** can denote the space of representations **Rep**, and in a sense **SoA** as well, since formal transformations also correspond to physical transformations. Nonetheless, if I am a Galilean, I will not admit any interpretation $\iota : \mathbf{Rep} \to \mathbf{SoA}$ that requires indicating a rest-frame to define. The result is that every admissible interpretation must associate all formal models that are related to one another by rest-frame transformation to the same state of affairs. This will yield a diagram with structure category $S_D = \mathbf{Gal}$. In other words, even though I take **Rep** and **SoA** to be indicated by **Gal***, my admissible interpretations are all mediated through **Gal** in the sense indicated by the Ockham diagram.

The answer to "what's so special about empty spacetime?" for expressing the concept of Galilean spacetimes is merely that **Gal** happens to correspond to the structure category S_D for (I expect) any reasonable understanding of how the Galilean formalism purports to latch on to possible states of affairs. In particular, the claim is not that **Gal** should be thought of as playing the role of either **Rep** or **SoA**. This lets me acknowledge the fact that rest-frame shifting isomorphisms can play a role in the theory, for both mathematical and physical possibilities. Rather, the claim is that these distinctions vanish when one considers the broader context of how a Galilean intends **Rep** to latch onto **SoA**. Both the event planner and their assistant can understand the chairs and the squares to be physically and symbolically "orientable", without thereby taking figure 1.1 to indicate a particular orientation.

1.4 Properties, Structure, and Stuff

The TCM procedure for determining the relative structural content of two theoretical formalisms T_1 and T_2, as employed in chapters 2 and 3, goes roughly as follows.

We begin with a question of the form:

(Q) Does T_1 have [more/less/the same] structure relative to T_2?

Answering this question involves a step by step process along the lines of:

(1) Formalism T_1 can be understood as category $\mathbf{C}_1$.

(2) T_2 can be understood as category $\mathbf{S}_2$.

(3) The relationship between T_1 and T_2 can be understood as a functor $F : \mathbf{S}_i \to \mathbf{S}_j$.[8]

(4) The functor F has the feature of (not) being [full/faithful/essentially surjective].[9]

From (4) is inferred that

(A) The physically relevant relationship between T_1 and T_2 is such that T_1 has [more/less/the same] [structure/properties/stuff] as T_2.

I have spent most of this chapter providing an account of how and why steps (1) and (2) make sense. That is, I have argued that a category $\mathbf{S}_i$ can be thought of as the *structure category* that summarizes ones prior commitments regarding a formalism's representational capacity. This section finishes the story by describing and motivating the "property-structure-stuff" (PSS) heuristic for comparing the relative content of formalisms.

[8]Or else the relationship is not functorial, in which case the theories are structurally incomparable, as in Barrett (2014).

[9]Perhaps along with a few other coherence criteria for the functor, such defineability (Hudetz, 2019) or non-splitting (Rosenstock and Weatherall, 2016).

1.4.1 Functor selection

Following the above inference schema, once we construct the structure categories $\mathbf{S}_1$ and $\mathbf{S}_2$ corresponding to formalisms T_1 and T_2, we proceed to construct a functor $F : \mathbf{S}_i \to \mathbf{S}_j$ that purports to summarize the relevant relationship between T_1 and T_2. But just as TCM arguments are vulnerable to accusations of arbitrariness in the choice of category to attach to each theoretical formalism, so too is it plagued by accusations concerning the functor used to compare the structural content of the categories chosen.

In section 1.3.3, we somewhat ameliorated the worry about arbitrariness of category choice by introducing the idea of a structure category $\mathbf{S}_D$, defined from a representation diagram $D = \mathbf{Rep} \overset{\vdots}{\underset{I}{\to}} \mathbf{SoA}$. D presents an externally sourced narrative about how a representation can be interpreted as physical states of affairs in a theoretical context, which S_D summarizes. Similarly, the functor $F : \mathbf{S}_i \to \mathbf{S}_j$ will serve to summarize an externally sourced story about how T_1 and T_2 relate to one another. This does not make the choice completely arbitrary, nor does it make TCM meaningless decoration. TCM provides a framework for precisely and succinctly expressing that relationship narrative in a way that is robust across insignificant representational choices, and can provide new insights into that relationship via the analysis presented in section 1.4.3.

Suppose I have a story about a sense in which T_2 can be thought of as having "more structure than" T_1. That is, I think that $T_2 = T_1+$ something more. Then I should be able to express T_1 using the the structural content $\mathbf{S}_2$ of T_2 as $I_1' : \mathbf{S}_2 \to \mathbf{SoA}_1$. The resulting Ockham diagram for any $\iota \in I_1'$ will induce a functor $F : \mathbf{S}_2 \to \mathbf{S}_1$. F is thus constructed in a unique way by attending to the sense in which $T_2 = T_1+$ something more.

1.4.2 Types of functors

Recall that a *functor* is a map $F : \mathbf{C}_1 \to \mathbf{C}_2$ between categories that plays nicely with morphisms. Below I describe three simple properties a functor can have, which we will use to analyze the relationship between the structure categories it connects.

F is said to be *essentially surjective* if for every object $y \in \text{Obj}(\mathbf{C}_2)$, there is an object $x \in \text{Obj}(\mathbf{C}_1)$ such that $F(x)$ is isomorphic to y in $\mathbf{C}_2$—that is, $\text{Hom}(\mathbf{C}_2)$ includes an isomorphism $h : F(x) \to y$. This "up to isomorphism" caveat is ubiquitous in (and largely characteristic of) category theory. When we have an isomorphism in a category, we are in a sense thinking of the objects it relates as the same. So for F to be *essentially* surjective is to say that every object in $\mathbf{C}_2$ is *essentially the same as* an object in the image of F. In the TCM context, this means that every theoretical model in $\mathbf{C}_1$ can be though of as effectively corresponding to some theoretical model in $\mathbf{C}_2$.

Similarly, F is *full* if for any objects $x, x' \in \mathbf{C}_1$, every morphism $h : F(x) \to F(x')$ in $\mathbf{C}_2$ has at least one corresponding morphism $h' \in \mathbf{C}_1$ such that $F(h') = h$. In the TCM context, this means that $\mathbf{C}_1$ can express at least as many relations between theoretical models x and x' as could be expressed by moving to $\mathbf{C}_2$ via F.

We say F is *faithful* if for any objects $x, x' \in \mathbf{C}_1$, no two distinct morphisms $g, h : x \to x'$ are such that $F(g) = F(h)$. This means that every morphism from x to y has a representative in the image of F in $\mathbf{C}_1$. This is in a sense the opposite of fullness—it means that $\mathbf{C}_2$ can express all of the relations between theoretical models x and x' that $\mathbf{C}_1$ can.

If a functor is full, faithful, and essentially surjective, then it realizes an *equivalence of categories*.

1.4.3 Mathematical gadgets

The analysis of functors used in this text is a byproduct of an email conversation in the late 1990s between mathematician John Baez and his graduate students Toby Bartels and James Dolan, archived in (Baez et al., 2013). The "properties-structure-stuff" taxonomy is rooted in the intuition that one can characterize what they refer to as "mathematical gadgets" as follows. One starts with some *stuff*—a set or collection of sets, an abstract "space", etc. This stuff can be equipped with *structure* in the form of special functions, relations, elements, subsets, etc. This "structured stuff" can also be conceived as satisfying *properties*, such as equations, inequalities, inclusions, etc. For example, one can express a *function* as

a pair of sets X, Y	*(stuff)*
equipped with $f \subseteq X \times Y$	*(structure)*
satisfying $\forall x \in X \quad \exists! y \in Y$ s.t. $(x, y) \in f$.	*(property)*

This provides a rough guide for how to construct a category associated with a mathematical gadget. One starts by specifying the objects of a category as structured stuff that satisfies a certain property, and then adds in the *morphisms* as maps from stuff to stuff that preserve the specified structure.

Now consider a functor $F : \mathbf{C}_1 \to \mathbf{C}_2$ between two categories defined in this way to capture mathematical gadgets G_1 and G_2. If the relationship between the gadgets that this functor encodes involves *forgetting a property*, then we should expect the functor to fail to be essentially surjective. That is, $\mathbf{C}_2$ would have all of the "structured stuff" that $\mathbf{C}_1$ has, plus more structured stuff that do not satisfy the property operative in G_1.

Similarly, suppose the relationship encoded by the functor involves *forgetting structure*. That is, we conceive of G_1 as possessing some additional structure relative to G_2. Then we should expect $F : \mathbf{C}_1 \to \mathbf{C}_2$ to fail to be full. This is because morphisms in $\mathbf{C}_1$ will be required to preserve that structure, while morphisms in $\mathbf{C}_2$ will not. So we should expect there to be extra morphisms in $\mathbf{C}_2$ that do not preserve the structure that G_1 has but G_2 does not.

Lastly, suppose F involves forgetting some stuff—for example, G_1 may be built from 2 sets, whereas G_2 involves only one of these. Then we would expect $F : \mathbf{C}_1 \to \mathbf{C}_2$ to fail to be faithful. That is, morphisms that behave identically on the first set but differently on the second in $\mathbf{C}_1$ will be mapped to the same morphism in $\mathbf{C}_2$.

1.4.4 Example: categories of squares

Let's start by considering the gadget "square", characterized by the category **Sq** whose objects are the collection of all squares, and whose morphisms consist of all of the ways you can transform one square into another. For example, you can make a square bigger, change its color, rotate it 90°, etc.

Now consider the gadget—"green square"—associated with the category **GrSq** whose objects are the collection of all green squares, and whose morphisms consist of all of the ways you can transform one green square into another. **GrSq** can naturally be thought of as a subcategory of **Sq** by an inclusion functor $F_1 : \mathbf{GrSq} \to \mathbf{Sq}$ that effectively acts as the identity.

Perhaps surprisingly, F_1 is an equivalence of categories! It is clearly full and faithful, since it is an inclusion. It is moreover essentially surjective, since every square is isomorphic to a green square via a color transformation. So F_1 forgets nothing, and conceived in this way, squares are equivalent to green squares.

In order to get the expected answer—that greens squares are squares + the property of greenness—one has to reify color as a meaningful feature of the squares. That is, consider the gadget "colored square" rather than the gadget "square". This can be associated with the category **ColSq** which has the same objects as **Sq**, but unlike in **Sq**, the morphisms of **ColSq** are required to preserve color. Now the inclusion functor F_2 : **GrSq** $\to$ **ColSq** fails to be essentially surjective, and so can be thought of as "forgetting properties". F_2 allows us to think of **GrSq** as the subcategory of **ColSq** that one obtains by requiring that objects satisfy the additional property of being green.

The sense in which **Sq** is equivalent to **GrSq** reflects the ambiguity with which **Sq** was defined in the first place. **Sq** is supposed to consist of "all squares"—but it is not practical, and arguably not possible, to enumerate all squares. But by requiring that all transformations that preserve squareness act as isomorphisms in our category, we can sidestep this issue. No matter which squares I think to enumerate, the resulting category I generate is "naturally" equivalent. This is what makes category theory so powerful, as it track concepts rather than particular ways of writing them out. It is only by adding or removing morphisms from these categories that I obtain a distinct gadget. This is how **ColSq** is defined—by "removing" color changing morphisms from **Sq**.

Now consider the gadget "smiley square" consisting of a square with a smiley face drawn on it. This is associated with the category **SmSq** whose objects are squares with smiley faces on them, and whose morphisms are those morphisms of the underlying square which moreover preserve the smiley face. Let F_3 : **SmSq** $\to$ **Sq** be the functor that takes smiley squares to their underlying squares. This functor is essentially surjective—every square in **Sq** can be thought of as the product of "removing" the smiley face from a square in **SmSq**. F_3 is also faithful, since every distinct smiley square transformation is also a distinct transformation of the underlying square. But F_3 is not full—there are transformations of squares in **Sq**—like being flipped upside-down—that do not appear in **SmSq** since they would not preserve the smiley face. Thus we can conceive of F_3 as "forgetting the structure" of the smiley face.

Lastly, consider the gadget "square + triangle" associated with the category **TrSq** whose objects are pairs consisting of a square and a triangle, and whose morphisms take squares to squares and triangles to triangles. Then the functor $F_4 : \mathbf{TrSq} \to \mathbf{Sq}$ that maps each pair to its square *forgets stuff*. It is essentially surjective, since every square can be thought of as half of a square + triangle pair. F_4 is full since every transformation of a square can be associated with the square-transforming component of a morphism in **TrSq**. But it is not faithful, since **TrSq** morphisms that act the same on their squares but differently on their triangles (e.g. by rotating them) will nonetheless map to the same square morphism in **Sq**.

However, there is also a "natural" functor going in the other direction that forgets structure! Let $F_5 : \mathbf{Sq} \to \mathbf{TrSq}$ that pairs each square with the same triangle "$\triangle$". F_5 is essentially surjective—every square + triangle pair can be arrived at by an isomorphic transformation of $\triangle$. It is also faithful—every square morphism corresponds to exactly one morphism of the pair consisting of that square + $\triangle$. But it fails to be full—morphisms in **TrSq** that rotate $\triangle$ do not have a preimage in **Sq**. The structure forgotten by F_5 can be thought of as fixing a representative $\triangle$ in a particular orientation.

Which captures the "true" relationships between the gadgets "square" and "square + triangle"? Neither is intrinsically "right", but there may be reasons to prefer one to the other based on the story one is trying to tell about **Sq** and **TrSq** in their capacity as denoting representational structure categories.

If I start with a gadget "square", then I can construct the gadget "square + triangle" by *adding* a triangle (stuff). This relationship is captured functor $F_4 : \mathbf{TrSq} \to \mathbf{Sq}$. But if I start with the gadget "square + triangle", then there is more than one way to connect it to the gadget square "square" via a functor $F : \mathbf{Sq} \to \mathbf{TrSq}$, corresponding to different choices of $\triangle$. To choose one of these amounts to "fixing" the triangle as it is, so that I am no longer free to move between triangles. This choice of triangle representative can thus be thought of extra structure. Pinning down a triangle removes a degree of freedom from my ability to use "square + triangle" to represent things in the world.

Let's look at this in terms of TCM. Consider two representation diagrams, D and $D_\triangle$, where $D = \mathbf{Rep} \overset{\vdots}{\underset{I}{\to}} \mathbf{SoA}$ has structure category $\mathbf{Sq}$, and $D_\triangle = \mathbf{Rep}_\triangle \overset{\vdots}{\underset{I_\triangle}{\longrightarrow}} \mathbf{SoA}_\triangle$ has structure category $\mathbf{TrSq}$. Suppose I take D to denote how I am using the gadget "square" to represent states of affairs. Since $\mathbf{Sq}$ is the structure category of D, then every $\iota \in I$ can be written as $\kappa_\iota \circ F_D$, in accordance with its Ockham diagram. Similarly, every $\iota_\triangle \in I_\triangle$ can be written as $\kappa'_{\iota_\triangle} \circ F_\triangle$. F_4 allows "reconstrue" D to use $\mathbf{Rep}_\triangle$ for its representations by creating a new representation diagram $D' = \mathbf{Rep}_\triangle \overset{\vdots}{\underset{I'_\triangle}{\longrightarrow}} \mathbf{SoA}$, where I reconstrue every $\iota = \kappa_\iota \circ F_D \in I$ as $\iota' = \kappa_\iota \circ F_4 \circ F_\triangle \in \mathbf{Rep}_\triangle$. Conversely, I can reconstrue $D_\triangle$ to use the representation space of $\mathbf{Rep}$ as $D'_\triangle = \mathbf{Rep} \overset{\vdots}{\underset{I'_\triangle}{\longrightarrow}} \mathbf{SoA}_\triangle$ with interpretations $\iota'_\triangle = \kappa'_{\iota_\triangle} \circ F_5 \circ F_D$.

In this context, F_4 might better capture the relationship between $\mathbf{Sq}$ and $\mathbf{TrSq}$ for someone who initially is employing the representation diagram D. That is, from the perspective of someone employing $\mathbf{Sq}$ to represent the world, $\mathbf{TrSq}$ appears to "add stuff" to my representations. Conversely, someone starting with an understanding of the world mediated by structure category $\mathbf{TrSq}$ might instead view the move from $\mathbf{TrSq}$ to $\mathbf{Sq}$ as adding the "structure" that fixes the triangle representative and orientation. To someone with no preconceptions about which gadget is more fundamental, both F_4 and F_5—and their accompanying stories about "squares" and "squares + triangles"—deserve consideration.

Because there can be multiple natural functors between gadget categories, one has to be careful about reading too much into relationships given by a particular functors. In particular, when presenting a functor $F : \mathbf{C}_1 \to \mathbf{C}_2$ as indicating that $\mathbf{C}_1$ has more structure or stuff than $\mathbf{C}_2$, one should check (as I do in chapter 3) whether F *splits*. That is, whether there exists a functor G going the other direction that composes with F to act as the identity. The absence of such a splitting functor indicates that the PSS verdict about F is more binding. The presence of one requires one to put both "stories" on equal footing as ways to compare the gadgets each category encodes, or else bring additional extra-formal considerations about the gadgets to bear on deciding in favor of one or another.

In his 2015 book *Ockham's Razors*, Elliott Sober considers the case of deciding between whether to employ linear or parabolic models to represent the relationship between variables x and y in a physical system. That is, a modeler is faced with two formalisms they might employ. They can use a gadget of the form $y = a_0 + a_1 x$ for real numbers a_0 and a_1, associated with a category **Lin**, or one of the form $y = a_0 + a_1 x + a_2 x^2$, associated with the category **Par**. The morphisms for both categories would be linear and parabolic transformations, respectively.

Sober explains how this choice poses a puzzle for philosophers of science. On the one hand, linear models are just a special case of parabolic models, so assuming a parabolic relationship between x and y is less presumptuous. On the other hand, if both work equally well, modelers almost universally prefer a linear framework, since it is requires assuming the existence of fewer constants. This ambivalence can be understood in terms of two different functors between **Lin** and **Par**. **Par** has *more stuff* than **Lin** in the sense that the functor from **Par** to **Lin** that "forgets" a_2 fails to be faithful. But **Par** also has *fewer properties* than **Lin** in the sense that the inclusion functor **Lin** $\hookrightarrow$ **Par** forgets the property that $a_2 = 0$.

1.4.5 Functor relativity

Talk about properties, structure, and stuff is suggestive of a characterization of individual formalisms. That was the motivation, after all—characterizing formalisms as stuff with structure and properties. Indeed, this way of thinking can be useful for constructing the categories to plug into TCM. However, the PSS heuristic is fundamentally attached to functors, not categories themselves. This makes the TCM notion of property, structure, and stuff fundamentally relational—a feature of relating two formalisms to one another. It is moreover attributed to particular way of relating formalisms using a functor, when there may be multiple viable options.

When considering a single formal gadget, one can "carve it up" in different ways into property + structure + stuff. For example, one can think of a metric space as a set (stuff) + a relation between elements (structure) that satisfies the property of being a "distance" relation. Alternatively, metric spaces can be characterized as a set (stuff) + a structure (distance relation) with no additional properties. Or, one could think of its stuff as being a topological space (since all metric spaces are topological spaces) with the property of being metrizable. Or one could take metric spaces to be primitive stuff, with no additional structure or properties.

Once a functor $F : \mathbf{C}_1 \to \mathbf{C}_2$ is chosen to characterize the relationship between two gadgets G_1 and G_2, it essentially treats the objects of $\mathbf{C}_2$ category as the "stuff" one builds the G_1s out of. It also treats the morphisms in $\mathbf{C}_2$ as a guide to the structure of G_1. The structure of G_1 is indicated by which morphisms from $\mathbf{C}_2$ do not have a preimage in $\mathbf{C}_1$. The properties of G_1 are defined in contrast to what would have been possible attributes of G_1 in the absence of those properties, as indicated by objects in $\mathbf{C}_2$ that do not have a preimage in $\mathbf{C}_1$.

For example, thinking of "greenness" as a "property" of the gadget "green square" implicitly invokes an inclusion functor $\mathbf{GrSq} \hookrightarrow \mathbf{ColSq}$ that forgets this property. But the inclusion $\mathbf{GrSq} \hookrightarrow \mathbf{Sq}$ is an equivalence, corresponding to conceiving of green squares as primitive stuff rather than squares (stuff) + greenness (property).

1.5 Some Caveats

The tools we have developed cannot serve as a guide to how to carve up a gadget G_1 into stuff + structure + properties. But in the giving of a functor $F : \mathbf{C}_1 \to \mathbf{C}_2$ characterizing a relationship between gadgets G_1 and G_2, we can define a way to carve up G_1 relative to this functor. How much significance should we read into a PSS story about G_1 so obtained, of the form "G_1 has more [structure/properties/stuff] than G_2"? The terms "property", "structure", and "stuff" are certainly suggestive of ontologically meaningful features of theoretical formalisms. The examples discussed in this chapter make such an interpretation promising, but it is unclear how far that generalizes.

The PSS story seems to latch onto features of the F-relationship between G_1 and G_2 that signal their relative representational capacity. When F forgets properties, then a move from understanding states of affairs as G_1 models to G_2 models via F essentially expands the scope of the representational capacity of G_1. F connects G_1 to G_2 in a way that allows us to think of G_2 as encompassing more possibilities that G_1—ways of being that do not satisfy a certain property that constraints G_1. When F forgets stuff, this move amounts to reducing the dimensionality of G_1 models—eliminating a degree of freedom in representation. When F forgets structure, it "smooths out" features of G_1, so that they no longer have representational significance.

These types of transformations—expanding scope, reducing dimension, increasing granularity—are all significant to modelers. Insofar as the PSS heuristic latches onto these sorts of moves, it does a good job of capturing the relevant relationships between formalisms. But that does not make applying and interpreting the PSS heuristic straightforward. I have already discussed that the story one gets from PSS is relative to a particular choice of functor, and beyond the challenge of motivating the construction of categories and functors, there may be many functors that are equally compelling. Additional challenges arise for PSS storytelling when a functor forgets more that one of properties, structure, and stuff.

For example, consider the gadgets "square" and "colored square". Since **Sq** is categorically equivalent to **GrSq**, there is a sense in which "colored square" forgets properties, given by the embedding of **Sq** into the green squares in **ColSq**. But it is not accurate to the intentions of modelers to say the PSS resulting story—that "squares" are "colored squares" + the property of greenness. A more promising relationship is given by the functor $F : \mathbf{ColSq} \to \mathbf{Sq}$ that “forgets” that we’re supposed to care about color, and identifies squares that are related to one another by a color change. This functor forgets stuff *and* structure. So even relative to this particular choice of functor, the relationship between squares and colored squares can be conceived of in two ways. One can think of colored squares as pairs consisting of a square and its color, so that F forgets color stuff. From this perspective, color is a degree of freedom available to modelers using colored squares, which is removed by instead using mere “squares”. Or one could think of colored squares as squares with “color structure”, so that F essentially “smooths out” this structure and presents squares more granularly. The act of removing a degree of freedom and of smoothing out a structure, while encoded by the same functor, imply different intentions on the part of a modeler making the modeling choice to move from colored squares to squares. This distinction cannot be fully captured by employing the PSS heuristic, but requires an additional narrative to accompany the category theoretic presentation.

The PSS story can also be misleading. Consider again the relationship between linear and parabolic models. We’ve discussed the senses in which parabolic models have more stuff than linear models, but fewer properties. But neither of the functors discussed fully account for what happens when a modeler transitions between linear and parabolic modeling paradigms. When transitioning from a parabolic to a linear framework, a modeler does not merely “drop” the a_2 term, and when transitioning from a linear to a parabolic framework, she does not keep the same model. Rather, in each framework, a particular linear or parabolic model is chosen because it fits the best with the data. While each formalism does not itself allow one to distinguish between states of affairs that have the same “best fitting curve”, they disagree with one another about how states of affairs should be “carved up” into equivalence classes that share a curve fit. So while I can represent the same underlying state of affairs (bare data sets) in either framework, their representation diagrams involve different states of affairs categories. I say that the PSS story is misleading in this case because there are nonetheless seemingly natural functors between **Lin** and **Par**, which might incline one towards a hasty

PSS verdict without carefully drawing out representation diagrams. I suspect that TCM could be useful for addressing this case nonetheless, though it would require analysis beyond simple PSS.

1.5.1 Will groupoids suffice?

While I have discussed TCM in terms of the general categories associated with theoretical formalisms, we have so far only invoked *groupoids*—categories in which all arrows are isomorphisms. Indeed, the TCM literature, including the examples discussed in chapters 2 and 3, only consider groupoids. But in motivating TCM, I highlighted the value of including all sorts of relations in the presentation of a formalism, including system-subsystem relations and projections onto lower dimensional representations. I do think there is good reason to focus on groupoids, but I also believe that such a focus fails to take advantage of all TCM has to offer philosophers of science, as I discuss in chapter 4.

Baez et al. motivate the focus on groupoids by noting that groupoids are less sensitive to arbitrary choices made in writing down a formalism. As an example, they discuss the following two PSS decompositions of a "monoid" in abstract algebra.

	stuff $\to$	a set M
monoid$_1$:	structure $\to$	a function $\cdot : M \times M \to M$, and an element $1 \in M$
	properties $\to$	$\forall x, y, z \in M, \quad (xy)z = x(yz)$ & $1x = x = x1$

	stuff $\to$	a set M
monoid$_2$:	structure $\to$	a function $\cdot : M \times M \to M$
	properties $\to$	$\forall x, y, z \in M, \quad (xy)z = x(yz),$ & $\exists e \in M \; \forall m \in M \quad em = m = me$
		(note that e is automatically unique.)

Recall that the PSS heuristic is motivated by understanding "morphisms" as maps from stuff to corresponding stuff that preserve structure.

For a monoid_1, a morphism is:

- a function $f : M \to M'$
- preserving multiplication and the unit: $f(xy) = f(x)f(y)$ and $f(1) = 1'$.

For a monoid_2, a morphism is:

- a function $f : M \to M'$
- preserving multiplication: $f(xy) = f(x)f(y)$.

These are really different, but they are the same in the case of *isomorphisms*. If f is an isomorphism of the first sort, it is obviously one of the second sort, but the converse holds too: if $f : M \to M'$ preserves multiplication, it preserves the unit too:

$$\begin{aligned} f(1) = 1' &\iff \forall y \in M', \quad f(1)y = y = f(1) \\ &\iff \forall y \in M', \quad f^{-1}(f(1)y) = f^{-1}(y) = f^{-1}(yf(1)) \\ &\iff \forall y \in M', \quad 1f^{-1}(y) = f^{-1}(y) = f^{-1}(y)1 \\ &\iff \forall z \in M, \quad 1z = z = z1. \quad \square \end{aligned}$$

However, there is no contradiction if an arbitrary (non-isomorphism) monoid_2 morphism f is such that $f(1) \neq 1'$. For example, if $M = M' = 2 \times 2$ matrices, the constant function

$f \equiv \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ is a monoid$_2$ morphism, even though $f \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \neq \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

The moral is that whether we count something as a structure or a property does not affect the *iso*morphisms, so the *groupoid* of a gadget is more robust than the category. This is also true about stuff that can be reinterpreted as structure (e.g., the unit of a monoid could be thought of as *stuff*: a one element set $\{*\}$ with the function $f : \{*\} \to M$).

In addition to groupoids being especially well-behaved, many other morphisms of interest in a sense "reduce" to isomorphisms. There are category theoretic analogs of the isomorphism theorems of abstract algebra that in many cases allow us to understand a given morphism in terms of isomorphisms between (possibly distinct) objects. (Mac Lane, 2013, chapter VIII)

1.6 A Modest Proposal

In light of the forgoing considerations, I suppose "theories as categories of models" is an inapt choice of designation after all. The phrasing may falsely suggest that a theory can be identified with a particular category. Not only do I want to allow for theories to involve physically meaningful content not captured by the formal structure of a category, I also believe it is often not justifiable to restrict attention to a single category when characterizing the role that a formalism plays in a theoretical context.

TCM is an attempt to characterize the structure of a theoretical formalism—the language-independent, essential content of a formal apparatus that is used to represent a physical system. A major barrier to doing this is the fact that the words we use to describe a theoretical formalism can be multiply interpreted, and often refer to genuinely different gadgets in a given theory depending on the particular context of application. The only sensible thing to do is try to figure out the functional role that a formalism is playing in various acts of theoretical representation. As I argue in section 1.3, this can be accomplished by constructing *representation diagrams* $D = \mathbf{Rep} \overset{\vdots}{\underset{I}{\to}} \mathbf{SoA}$. I show in section 1.3.3 that one can uniquely define a *structure category* S_D from a representation diagram, which summarizes the way in which interpretations in that diagram attach states of affairs to formal representations. While constructing D will involve countless arbitrary choices, insofar as various acts of representation can be said to employ the "same" formal structure, we can expect that the structure categories will coincide.

This resolution does not provide The True Category™ associated with a theoretical formalism on a platter from heaven. Rather, it shows how a particular category $\mathbf{S}_D$ can be constructed in a unique way from a sufficiently precise description of the manner in which a theory uses a formalism to interpret representations as states of affairs. In other words, the starting place for understanding theories as categories of models is a pre-existing narrative about how formalisms are used to represent physical systems. This narrative can be derived from observations, experience, intuitions, or any number of sources. TCM is silent as to which narrative is *preferable*; it merely amounts to imposing additional specificity and consistency constraints to make the narrative analytically tractable. But once we have presented theories as categories of models, we have access to additional tools from category theory, such as the PSS heuristic, that we can employ to assist in making decisions as to which narrative is indeed preferable.

The label "theories as categories of models" should perhaps instead be thought of as shorthand for "the structure categories of representation diagrams are useful for pinning down the intended structural content a theoretical formalism as employed in a given act of representation." The rest of this text will give explicit examples of how it can be put to use.

Chapter 2

General Relativity

The main result in this chapter was achieved in collaboration with Thomas Barrett and James Weatherall, now published in (Rosenstock et al., 2015).

2.1 Introduction

In chapter 1, I discussed how Leibniz viewed the concept of "absolute rest" in Newtonian spacetime theories as unwarranted additional structure on spacetime, not necessary to account for the empirical adequacy of Newton's theory. The act of eliminating this excess structure can be characterized by a functor $F : \mathbf{Newt} \to \mathbf{Gal}$ that "forgets" the rest frame structure. Spacetime theories have progressed since Newtons time, but the urge to continue whittling down "excess structure" in order to reveal the fundamental nature of spacetime has persisted among physicists and philosophers. The rest of this text will demonstrate how TCM can be similarly invoked to address debates about the structural content of modern physical theories.

In this chapter, I apply the TCM framework to the question of whether the structure of a "spacetime substance" can be eliminated from general relativity to yield a more parsimonious formulation. This work responds to John Earman's (1986) proposal that Robert Geroch's (1972) "Einstein algebra" formalism for general relativity might do the trick. To address this question, I cast general relativity and the theory of Einstein Algebras as the categories **GR** and **EA** using the TCM framework outlined in the previous chapter. I present a functor that I argue captures the sense in which Einstein algebras represent the same states of affairs as general relativity, and prove that relative to this functor, **EA** and **GR** are equivalent. The functor that realizes the equivalence has the interesting property of being *contraviariant*, meaning that it "flips" homomorphisms. Such a functor is sometimes referred to as realizing *duality* rather than an equivalence, but the general TCM narrative about equivalence still holds.

I begin in section 2.2 by situating this project in the context of the relationism-substantivalism debate in philosophy of physics that inspired it. Section 2.3 introduces the concept of a *smooth algebra* following mathematician Jet Nestruev (2006), and demonstrates a categorical equivalence between the category **SmoothAlg** associated with smooth algebras and the category **SmoothMan** associated with more familiar formalism of smooth manifolds. Section 2.4 formally defines an Einstein algebra, and proves an analogous equivalence between this formalism and that of a manifold with metric—the mathematical object traditionally associated with spacetime in general relativity. I conclude in section 2.5 by discussing the implications of this result.

2.2 Relationism vs. substantivalism

Among philosophers of physics, questions about the content of spacetime theories are often cast in terms of the debate between relationist and substantivalist views.[1] The designations “relationism” and “substantivalism” each refer to a broad class of views that can be understood roughly as follows. *Relationists* understand spacetime to be in some sense an emergent property of matter and the relations between material objects, rather than a thing in itself. *Substantivalists* conversely attribute some degree of reality to spacetime independent of any matter present. This disagreement appears to entail meaningful differences in beliefs about the nature of the universe, rather than merely being different ways of describing the same phenomena. Relationists and substantivalists disagree about the counterfactual claim regarding whether spacetime would exist at all without matter.

This debate has been a fixture in philosophy since antiquity, and plays a central role in Newton and Leibniz’s disagreement about the nature of spacetime. A spacetime substance, if present, might motivate setting a standard of rest in Newtonian spacetime—namely, velocity could be thought of as defined relative to that fixed substance. Substantivalist considerations similarly motivated late nineteenth century aether theories, which purported to provide a standard of rest to justify the constant speed of light in Maxwell’s equations for electromagnetism. Relationist considerations played a major role in the subsequent development of general relativity (GR). Relationism appears somewhat explicitly in GR in the form of the *principle of relativity*, which requires that admissible dynamics cannot depend on a particular frame of reference. GR has been widely accepted by physicists to be the most empirically adequate spacetime theory to date for almost a century now (modulo some accommodations to allow for quantum phenomena). But the broad acceptance of GR has not fully alleviated relationist-substantivalist tensions.

[1] See (Huggett and Hoefer, 2018) for a historical overview.

For some relationists, the inclusion of a *spacetime manifold* in standard formulations of GR bears the stain of lingering substantival commitments. The spacetime manifold is the geometric object on which matter fields are defined, and is often colloquially referred to as just "spacetime". To modern-day relationists, belief in the truth of GR then troublingly seems to require commitment to this "spacetime" being real in some way.

It is in this context that Earman (1986) proposes that the standard formalism of general relativity be replaced with one that instead uses *Einstein algebras*, which he claims possesses all of the virtues of GR but without requiring a spacetime manifold. Instead, Einstein algebras express the relations among possible configurations of matter in an abstract way. Thus, Earman argues that the theory of Einstein algebras may be a fully relational spacetime theory, strictly preferable to the traditional conception of GR.

This is exactly the sort of situation TCM has been developed to analyze. There are two formalisms—manifolds with metric and Einstein algebras—which can be used to represent the same class of physical system: the universe according to general relativity. Earman has claimed that one of these eliminates superfluous structure from the other. That is, Einstein algebras are alleged to be more parsimonious, and less metaphysically presumptuous, than manifolds with metric. This chapter defines categories associated to each of these formalisms, and establishes a canonical functorial relationship between them.

Manifolds with metric—or relativistic spacetimes, as we will call them, can be straightforwardly understood as a category **GR**. The morphisms of *GR* are *isometric embeddings*, which are the accepted standard morphism for relativistic spacetimes. Einstein algebras, on the other hand, have not been widely employed by physicists, so there is no universally accepted notion of morphism for this object. Much of this text will be devoted to defining and motivating "Einstein algebra morphisms" in order to establish the category **EA** of Einstein algebras.

Section 2.3 provides the necessary background for this endeavor by reviewing the concept of a *smooth algebra* that underlies an Einstein algebra. It will be shown that the category **SmoothAlg** of smooth algebras is dual to the category **SmoothMan** of smooth manifolds. The duality between **EA** and **GR** shown in section 2.4 bootstraps off of this underlying duality between smooth algebras and smooth manifolds.

2.3 Smooth algebras and smooth manifolds

In what follows, the term *algebra* refers to a commutative, associative algebra with unit over $\mathbb{R}$—i.e., a real vector space endowed with a commutative, associative product and containing a multiplicative identity.[2] An *(algebra) homomorphism* is a map that preserves the vector space operations, the product, and the multiplicative identity; a bijective algebra homomorphism is an *(algebra) isomorphism*.

2.3.1 Smooth algebras

Let A be an algebra. $|A|$ denotes the collection of homomorphisms from A to $\mathbb{R}$. The elements of $|A|$ are known as the *points* of the algebra A; $|A|$ itself is the *dual space of points*.[3] (Note, however, that no algebraic structure is imposed on $|A|$.) An algebra A is *geometric* if there

[2]The treatment of smooth algebras in this text follows Nestruev (2006). For more on Einstein algebras in particular, see Geroch (1972) and Heller (1992).

[3]In some treatments of related material, including Rynasiewicz (1992), "points" are reconstructed as maximal ideals of appropriate rings. The present approach emphasizes the sense in which points are "dual" to smooth functions in the same sense of duality that one encounters elsewhere in geometry and algebra. But it is closely related to the approach Rynasiewicz (1992) uses. In particular, if x is an element of $|A|$, then $\ker(x)$ is an ideal, since if $f \in \ker(x)$, then for any $g \in A$, $x(fg) = x(f)x(g) = 0$; moreover, it is maximal, since by linearity, x is surjective, and thus $A/\ker(x) = \mathbb{R}$ and it is well known that for a commutative, unital ring (or algebra) A, an ideal I is maximal if and only if A/I is a field. Conversely, as we note above, if A is geometric, then A is canonically isomorphic to the space $\tilde{A}$, the maximal ideals of which consist in all functions vanishing at a given point $x \in |A|$. So points in the sense that Ryansiewicz considers uniquely determine points in the present sense, and vice versa.

are no non-zero elements $f \in A$ that lie in the kernel of all of the elements of $|A|$, i.e., if $\bigcap_{p\in|A|} \ker(p) = \{\mathbf{0}\}$.[4]

The space $\tilde{A}$ is defined as as follows:

$$\tilde{A} = \{\tilde{f} : |A| \to \mathbb{R} : \exists f \in A \text{ s.t. } \tilde{f}(x) = x(f))\}.$$

There is a natural algebraic structure on $\tilde{A}$, with operations given by:

$$(\tilde{f} + \alpha\tilde{g})(x) = \tilde{f}(x) + \alpha\tilde{g}(x) = x(f) + \alpha x(g)$$
$$(\tilde{f} \cdot \tilde{g})(x) = \tilde{f}(x) \cdot \tilde{g}(x) = x(f) \cdot g(f)$$

There is a canonical map $\tau : A \to \tilde{A}$ defined by $f \mapsto \tilde{f}$. In general, τ is a surjective homomorphism. For geometric algebras, however, τ is also injective, and thus an isomorphism. This enables us to freely identify a geometric algebra A with $\tilde{A}$ through implicit appeal to τ.

Given a geometric algebra A, the *weak topology* on $|A|$ is the coarsest topology on $|A|$ relative to which every element of A (or really, $\tilde{A}$) is continuous. This defines a Hausdorff topology on A. Now suppose we have an algebra homomorphism $\psi : A \to B$. Then ψ determines a map $|\psi| : |B| \to |A|$ defined by $|\psi| : x \mapsto x \circ \psi$. Any map $|\psi|$ that arises this way is continuous in the weak topology; if ψ is an isomorphism, then $|\psi|$ is a homeomorphism.

Now let A be a geometric algebra, and suppose that $S \subseteq |A|$ is any subset of its dual space of points. Then the *restriction* $A_{|S}$ of A to S is the set of all functions $f : S \to \mathbb{R}$ such that for any point $x \in S$, there exists an open neighborhood $O \subseteq |A|$ containing x, and an

[4]The expression "geometric algebra" is also used (somewhat more often) to described so-called Clifford algebras. See, for instance, Hestenes and Sobczyk (2012) or Doran et al. (2003). The present sense of the term is unrelated.

element $\bar{f} \in A$ such that f and $\bar{f}$ agree on all points in O. One easily verifies that $A_{|S}$ is an algebra, though it is not in general a subalgebra of A.

Given any $S \subseteq |A|$, we can define a homomorphism $\rho_S : A \to A_{|S}$, defined by $f \mapsto f_{|S}$, where here the restriction $f_{|S}$ is meant in the ordinary sense. The map ρ_S is known as the *restriction homomorphism.* A special case of restriction is restriction to $|A|$, i.e., to the dual space of the algebra, $A_{||A|}$. This is the collection of all maps on A that are "locally equivalent" to elements of A. We will say that A is *complete* if it contains all maps of this form—i.e., if the restriction homomorphism $\rho_A : A \to A_{||A|}$ is surjective.

A complete, geometric algebra A is called *smooth* if there exists a finite or countable open covering $\{U_k\}$ of the dual space $|A|$ such that all the algebras $A_{|U_k}$ are isomorphic to the algebra $C^\infty(\mathbb{R}^n)$ of smooth functions on $\mathbb{R}^n$, for some fixed n. Here n is known as the *dimension* of the algebra. Note that this sense of dimension is unrelated to the dimension of the vector space underlying A.

2.3.2 The duality of smooth algebras and manifolds

Smooth algebras and smooth manifolds bear a close relationship to one another. In what follows, this relationship is presented category theoretically, largely following Nestruev's (2006) non-categorical presentation. First, we define two categories: the category **SmoothMan**, whose objects are smooth manifolds and whose arrows are smooth maps, and the category **SmoothAlg**, whose objects are smooth algebras and whose arrows are algebra homomorphisms. We show that these two categories are dual to one another. This result will be of crucial importance in our discussion of Einstein algebras and relativistic spacetimes.

There is a way to "translate" from the framework of smooth manifolds into the framework of smooth algebras. We call this translation F and define it as follows.

- Given a smooth manifold M, $F(M) = C^\infty(M)$ is the algebra of smooth scalar functions on M.

- Given a smooth map $\phi : M \to N$, $F(\phi)$ is the map $\hat{\phi} : C^\infty(N) \to C^\infty(M)$ given by $\hat{\phi}(f) = f \circ \phi$ for any $f \in C^\infty(N)$.

Before showing that it is a contravariant functor between **SmoothMan** and **SmoothAlg**, let us notice a few features of F. Let M be a manifold and consider the algebra $F(M) = C^\infty(M)$ of smooth scalar functions on M. There is a natural correspondence between points in M and elements of $|C^\infty(M)|$, given by the following map:

$$\theta_M : M \to |C^\infty(M)| \qquad \theta_M(p)(f) = f(p) \tag{2.1}$$

for any $p \in M$ and $f \in C^\infty(M)$. Note that $\theta_M(p)$ is indeed a homomorphism $C^\infty(M) \to \mathbb{R}$. One can easily verify that the algebra $C^\infty(M)$ is geometric, so we can consider the weak topology on $|C^\infty(M)|$. One then proves that relative to the weak topology the map $\theta_M : M \to |C^\infty(M)|$ is a homeomorphism (Nestruev, 2006, 7.4).

This fact allows one to prove the following simple result. The map F translates a smooth manifold into a smooth algebra.

Proposition 2.1. If M is a smooth manifold, then $F(M) = C^\infty(M)$ is a smooth algebra (Nestruev, 2006, 7.5–7.6).

The next important result about F captures a sense in which the smooth maps between manifolds are characterized purely by their action on the algebras of smooth scalar functions.

Proposition 2.2. Let M and N be smooth manifolds. A map $\phi : M \to N$ is smooth if and only if $\hat{\phi}(C^\infty(N)) \subset C^\infty(M)$ (Nestruev, 2006, 7.16–7.18).

These results allow one to make precise a sense in which F is indeed a "translation" from the framework of smooth manifolds into the framework of smooth algebras. This leads us to the following result.

Lemma 2.3. $F : \mathbf{SmoothMan} \to \mathbf{SmoothAlg}$ is a contravariant functor.

Proof. Proposition 2.1 immediately implies that $F(M)$ is indeed an object in **SmoothAlg** for every smooth manifold M. Let $\phi : M \to N$ be a smooth map. We need to show that the map $F(\phi) = \hat{\phi} : F(N) \to F(M)$ is an algebra homomorphism. Proposition 2.2 implies that $\hat{\phi}(f) \in F(M)$ for every $f \in F(N)$. One can easily verify that $\hat{\phi}$ preserves the vector space operations, the product, and the multiplicative identity, so $F(\phi) = \hat{\phi}$ is an algebra homomorphism. Furthermore, it is easy to see that F preserves identities and reverses composition. This implies that $F : \mathbf{SmoothMan} \to \mathbf{SmoothAlg}$ is a contravariant functor. □

There is also a way to "translate" from the framework of smooth algebras into the framework of smooth manifolds. In order to describe this translation we need to do some work. Let A be a smooth algebra. One can use the smooth algebraic structure of A to define a smooth manifold $G(A)$. The underlying point set of the manifold $G(A)$ is the set $|A|$ of points of the algebra A.

Since A is a smooth algebra, there is a covering of $|A|$ by open sets $\{U_k\}$ along with isomorphisms $i_k : A_{|U_k} \to C^\infty(\mathbb{R}^n)$ for some fixed n. These open sets and isomorphisms can be used to define charts (U, ψ) on $|A|$. First consider the maps

$$h_k : A \to C^\infty(\mathbb{R}^n) \qquad h_k = i_k \circ \rho_{U_k},$$

where $\rho_{U_k} : A \to A_{|U_k}$ is the restriction homomorphism. One can verify that the maps $|\rho_{U_k}| : |A_{|U_k}| \to U_k \subset |A|$ are homeomorphisms onto U_k (Nestruev, 2006, 7.7-7.8). Since i_k is an isomorphism and $|C^\infty(\mathbb{R}^n)| = \mathbb{R}^n$ (Nestruev, 2006, 3.16), it follows that $|h_k| = |\rho_{U_k}| \circ |i_k|$ is a homeomorphism $|h_k| : \mathbb{R}^n \to U_k$. This allows us to define the charts (U_k, ψ_k), where $\psi_k = |h_k|^{-1}$ for each $k \in \mathbb{N}$. One can verify that these charts are compatible (Nestruev, 2006, 7.10).

In addition to these charts (U_k, ψ_k), one can add charts of the form $(V \cap U_k, \psi_k)$ where $V \subset |A|$ is an open set and $k \in \mathbb{N}$. It is easily verifiable that these new charts are compatible both with each other and with the original charts (U_k, ψ_k). Since the topology on $|A|$ is Hausdorff and since there is a countable cover of charts of the form (U_k, ψ_k), if we throw in wholesale all of the charts on $|A|$ that are compatible with the charts of the form (U_k, ψ_k) and $(V \cap U_k, \psi_k)$, then we will have defined a smooth (Hausdorff, paracompact) manifold (Malament, 2012, Proposition 1.1.1). This smooth manifold is denoted as $G(A)$.

The smooth manifold $G(A)$ bears a close relationship to the original smooth algebra A. Indeed, the elements of A correspond to smooth scalar functions on $G(A)$. This correspondence is given by the following map:

$$\eta_A : A \to FG(A) \qquad \eta_A : f \longmapsto (p \mapsto p(f)) \tag{2.2}$$

for every $f \in A$ and $p \in G(A) = |A|$. One can prove that for every $f \in A$ the function $p \mapsto p(f)$ is a smooth scalar function on $G(A)$, and furthermore, that the map η_A is a bijection (Nestruev, 2006, 7.11). The elements of A can therefore be thought of as smooth scalar functions on the manifold $G(A)$.

The translation G from the framework of smooth algebras into the framework of smooth manifolds is defined as follows.

- Given a smooth algebra A, $G(A)$ is the smooth manifold defined above.

- Given an algebra homomorphism $\psi : A \to B$, $G(\psi)$ is the map $|\psi| : |B| \to |A|$ between the manifolds $G(B)$ and $G(A)$.

Note that the definition of $G(\psi)$ makes sense since $G(B)$ and $G(A)$ have underlying point sets $|A|$ and $|B|$, respectively. Like Lemma 2.3, the following result captures a sense in which G is a translation between these two frameworks.

Lemma 2.4. G : **SmoothAlg** $\to$ **SmoothMan** is a contravariant functor.

Proof. It has already been shown that $G(A)$ is a smooth manifold for every smooth algebra A. Let $\psi : A \to B$ be an algebra homomorphism. We need to show that $G(\psi) = |\psi| : |B| \to |A|$ is a smooth map between the manifolds $G(B)$ and $G(A)$. We begin by showing that

$$\widehat{|\psi|} \circ \eta_A = \eta_B \circ \psi \tag{2.3}$$

For every $f \in A$ and $p \in |B|$ we see that following equations hold:

$$\begin{aligned}(\widehat{|\psi|} \circ \eta_A(f))(p) &= (\eta_A(f) \circ |\psi|)(p) \\ &= \eta_A(f)(p \circ \psi) \\ &= (p \circ \psi)(f) \\ &= (\eta_B \circ \psi(f))(p)\end{aligned}$$

The first equality follows from the definition of $\widehat{|\psi|}$, the second from the definition of $|\psi|$, the third from the definition of η_A, and the fourth from the definition of η_B. This establishes equation (2.3). Since the maps η_A and η_B are bijections, (2.3) implies that $\widehat{|\psi|} = \eta_B \circ \psi \circ \eta_A^{-1}$. And this means that $\widehat{|\psi|}(FG(A)) \subset FG(B)$. Proposition 2.2 then guarantees that $|\psi| : G(B) \to G(A)$ is a smooth map. One easily verifies that G preserves identities and reverses composition, so G : **SmoothAlg** $\to$ **SmoothMan** is a contravariant functor. □

This demonstrates that the maps $F : \mathbf{SmoothMan} \to \mathbf{SmoothAlg}$ and $G : \mathbf{SmoothAlg} \to \mathbf{SmoothMan}$ are contravariant functors. They are also "up to isomorphism" inverses of one another. The following theorem makes this idea precise.

Theorem 2.5. The categories **SmoothMan** and **SmoothAlg** are dual.

Proof. We show that the families of maps $\eta : 1_{\mathbf{SmoothAlg}} \Rightarrow F \circ G$ and $\theta : 1_{\mathbf{SmoothMan}} \Rightarrow G \circ F$ defined in equations (2.1) and (2.2) are natural isomorphisms. Since F and G are contravariant functors, this will imply that **SmoothMan** and **SmoothAlg** are dual categories.

We first consider η. We need to verify that for every smooth algebra A the component $\eta_A : A \to FG(A)$ is an algebra isomorphism. We have already seen that η_A is bijective. One easily checks that η_A preserves the vector space operations, the product, and the multiplicative identity. Equation (2.3) implies that naturality square for η commutes, so $\eta : 1_{\mathbf{SmoothAlg}} \Rightarrow F \circ G$ is a natural isomorphism.

We now consider θ. We need to verify that for every smooth manifold M the component $\theta_M : M \to GF(M)$ is a diffeomorphism. We already know that it is bijective. We show that $\hat{\theta}_M(FGF(M)) \subset F(M)$ and then use Proposition 2.2 to conclude that θ_M is smooth. Let $f \in FGF(M)$. Since $\eta_{F(M)} : F(M) \to FGF(M)$ is a bijection there is some $g \in F(M)$ such that $\eta_{F(M)}(g) = f$. We then see that the following equalities hold for every $p \in M$:

$$\hat{\theta}_M(f)(p) = \hat{\theta}_M(\eta_{F(M)}(g))(p) = (\eta_{F(M)}(g) \circ \theta_M)(p) = \theta_M(p)(g) = g(p)$$

The first equality holds by our choice of the function $g \in F(M)$, the second by the definition of $\hat{\theta}_M$, the third by the definition of $\eta_{F(M)}$, and the fourth by the definition of θ_M. This implies that $\hat{\theta}_M(f) = g \in F(M)$. So we have shown that $\hat{\theta}_M(FGF(M)) \subset F(M)$, which by Proposition 2.2 means that $\theta_M : M \to GF(M)$ is a smooth map. One argues in an analogous manner to show that θ_M^{-1} is smooth. Therefore $\theta_M : M \to GF(M)$ is a diffeomorphism.

We also need to show that the naturality square for θ commutes. Let $\phi : M \to N$ be a smooth map. We show that $GF(\phi) \circ \theta_M = \theta_N \circ \phi$. For every $p \in M$ and $f \in F(N)$ we see that the following equalities hold:

$$\begin{aligned}(|\hat{\phi}| \circ \theta_M(p))(f) &= (\theta_M(p) \circ \hat{\phi})(f) \\ &= \theta_M(p)(f \circ \phi) \\ &= f \circ \phi(p) \\ &= (\theta_N \circ \phi(p))(f)\end{aligned}$$

The first equality follows from the definition of $|\hat{\phi}|$, the second from the definition of $\hat{\phi}$, the third from the definition of θ_M, and the fourth from the definition of θ_N. Since p and f were arbitrary, this implies that $GF(\phi) \circ \theta_M = \theta_N \circ \phi$, so $\theta : 1_{\mathbf{SmoothMan}} \Rightarrow G \circ F$ is a natural isomorphism, and the categories **SmoothMan** and **SmoothAlg** are dual. $\square$

Theorem 2.5 allows one to identify the smooth algebra A with $FG(A) = C^\infty(|A|)$ and the smooth manifold M with $GF(M) = |C^\infty(M)|$. In addition, one can identify an algebra homomorphism $\psi : A \to B$ between smooth algebras with $FG(\psi) = \widehat{|\psi|}$ and a smooth map $\phi : M \to N$ between smooth manifolds with $GF(\phi) = |\hat{\phi}|$. Now that they have been made precise, these identifications will be implicitly assumed in what follows.

2.4 Einstein algebras and relativistic spacetimes

The theory of Einstein algebras proceeds by taking a 4-dimensional smooth algebra A—which by Theorem 2.5 corresponds to some smooth 4-dimensional manifold—and defining additional structure on it. This structure corresponds to the various fields that one requires to formulate general relativity. This structure is defined as follows.

Let A be a smooth algebra. A *derivation on* A is an $\mathbb{R}$-linear map $\hat{X} : A \to A$ that satisfies the Leibniz rule, in the sense that

$$\hat{X}(fg) = f\hat{X}(g) + g\hat{X}(f)$$

for all $f, g \in A$. The space of all derivations on A is a module over A. The notation $\hat{\Gamma}(A)$ will denote this module and $\hat{\Gamma}^*(A)$ will denote the dual module. The elements of the dual module $\hat{\Gamma}^*(A)$ are just the A-linear maps $\hat{\Gamma}^*(A) \to A$. Derivations on A allow one to define an analog to "tangent spaces" on smooth algebras. Given a derivation $\hat{X}$ on A and a point $p \in |A|$, one can consider the linear map $\hat{X}_p : A \to \mathbb{R}$ defined by $\hat{X}_p(f) = \hat{X}(f)(p)$. The *tangent space to* A at a point $p \in |A|$ is the vector space T_pA whose elements are maps $\hat{X}_p : A \to \mathbb{R}$. The cotangent space to A at a point $p \in |A|$ is defined similarly.

Derivations $\hat{X}$ on the smooth algebra A naturally correspond to ordinary smooth vector fields X on the manifold $G(A) = |A|$. The correspondence is given by

$$\hat{X}(f)(p) = X_p(f) \tag{2.4}$$

where $f \in C^\infty(|A|) = A$ and $p \in |A|$. This correspondence plays a crucial role in the following results, so we take a moment here to unravel the idea behind it. Given a derivation $\hat{X}$ on A, equation (2.4) defines a vector field X on the manifold $G(A)$. This vector field X assigns the vector X_p to the point $p \in G(A)$, where the vector X_p is defined by its action $f \mapsto \hat{X}(f)(p)$ on smooth scalar functions $f \in A$ on the manifold $G(A)$. One uses the fact that $\hat{X}$ satisfies the Leibniz rule to show that X_p is indeed a vector at the point $p \in |A|$. One also verifies that the vector field X is smooth.

Conversely, given a vector field X on the manifold $G(A)$, equation (2.4) defines the derivation $\hat{X}$ on A. The derivation $\hat{X}$ maps an element $f \in A$ to the element of A defined by the scalar function $X(f)$ on the manifold $G(A)$. It follows immediately that $\hat{X}$ is linear and satisfies the Leibniz rule. One can argue in a perfectly analogous manner to show that elements of $\hat{\Gamma}^*(A)$ correspond to smooth covariant vector fields on the manifold $G(A)$. Note also that given a point $p \in |A|$ the correspondence (2.4) allows one to naturally identify the elements $\hat{X}_p$ of the tangent space T_pA and the vectors X_p at the point p in the manifold $G(A)$.

A *metric* on a smooth algebra A is a module isomorphism $\hat{g} : \hat{\Gamma}(A) \to \hat{\Gamma}^*(A)$ that is symmetric, in the sense that $\hat{g}(\hat{X})(\hat{Y}) = \hat{g}(\hat{Y})(\hat{X})$ for all derivations $\hat{X}$ and $\hat{Y}$ on A. A metric $\hat{g}$ on A induces a map $\hat{\Gamma}(A) \times \hat{\Gamma}(A) \to A$ defined by

$$\hat{X}, \hat{Y} \longmapsto \hat{g}(\hat{X})(\hat{Y})$$

Given a point $p \in |A|$, a metric on A also induces a map $T_pA \times T_pA \to \mathbb{R}$ defined by $\hat{X}_p, \hat{Y}_p \mapsto \hat{g}(\hat{X}, \hat{Y})(p)$. We will occasionally abuse notation and use $\hat{g}$ to refer to all three of these maps, but it will always be clear from context which map is intended.

If $\hat{g}$ is a metric on an n-dimensional smooth algebra A and p is a point in $|A|$, then there exists an m with $0 \leq m \leq n$ and a basis $\xi_1, \ldots, \xi_n$ for the tangent space T_pA such that

$$\begin{aligned}
\hat{g}(\xi_i, \xi_i) &= +1 \qquad \text{if} \quad 1 \leq i \leq m \\
\hat{g}(\xi_j, \xi_j) &= -1 \qquad \text{if} \quad m < j \leq n \\
\hat{g}(\xi_i, \xi_j) &= 0 \qquad \text{if} \quad i \neq j
\end{aligned}$$

The pair $(m, n-m)$ is called the *signature* of $\hat{g}$ at the point $p \in |A|$. A metric $\hat{g}$ on $|A|$ that has signature $(1, n-1)$ at every point $p \in |A|$ is called a metric of *Lorentz signature.*

This gives us the resources necessary to begin discussing the theory of Einstein algebras. An *Einstein algebra* is a pair $(A, \hat{g})$, where A is a smooth algebra and $\hat{g}$ is a metric on A of Lorentz signature. Before proving that the category of Einstein algebras is dual to the category of relativistic spacetimes, we need some basic facts about the relationship between metrics on algebras and metrics on manifolds.

Lemma 2.6. Let M be an n-dimensional smooth manifold and let A be an n-dimensional smooth algebra. Then the following all hold:

(1) If g is a Lorentzian metric on M, then $\hat{g}$ is a Lorentzian metric on the algebra $F(M) = C^\infty(M)$, where $\hat{g}(\hat{X})(\hat{Y}) := g(X, Y)$;

(2) If $\hat{g}$ is a Lorentzian metric on A, then $|\hat{g}|$ is a Lorentzian metric on the manifold $G(A) = |A|$, where $|\hat{g}|(X, Y) := \hat{g}(\hat{X})(\hat{Y})$;

(3) If g is a metric on a M, then $|\hat{g}| = g$;

(4) If $\hat{g}$ is a metric on A, then $\widehat{|\hat{g}|} = \hat{g}$.

Proof. Let g be a Lorentzian metric on M. One can easily verify that the map $\hat{g} : \hat{\Gamma}(F(M)) \to \hat{\Gamma}^*(F(M))$ defined by $\hat{g}(\hat{X})(\hat{Y}) = g(X, Y)$ is a symmetric module isomorphism, and therefore a metric on the smooth algebra $F(M) = C^\infty(M)$. It immediately follows from the bilinearity of g that $\hat{g}$ is a module homomorphism; that $\hat{g}$ is bijective and symmetric follows from the fact that g is non-degenerate and symmetric.

We also need to show that $\hat{g}$ has Lorentz signature. Let $p \in M$ and let $\xi_1, \ldots, \xi_n$ be an orthonormal basis (relative to the metric g) for the tangent space T_pM. Vectors at $p \in M$ can be naturally identified via (2.4) with elements of the tangent space $T_pF(M)$ to the algebra $F(M) = C^\infty(M)$. This identification and the definition of $\hat{g}$ immediately imply that $\hat{g}$ must have the same signature as g. So $\hat{g}$ is a metric of Lorentz signature on $F(M)$ and therefore (1) holds. One argues in an analogous manner to demonstrate (2).

If g is a metric on M and X and Y are vector fields on M, then $|\hat{g}|(X, Y) = \hat{g}(\hat{X})(\hat{Y}) = g(X, Y)$. Furthermore, if $\hat{g}$ is a metric on A and $\hat{X}$ and $\hat{Y}$ are derivations on A, then $\widehat{|\hat{g}|}(\hat{X})(\hat{Y}) = |\hat{g}|(X, Y) = \hat{g}(\hat{X})(\hat{Y})$. This immediately implies (3) and (4). □

Lemma 2.6 captures a sense in which metrics on manifolds and metrics on smooth algebras encode exactly the same information. Each kind of structure naturally induces the other. This lemma strongly suggests, therefore, that we will be able to recover a sense in which general relativity and the theory of Einstein algebras are equivalent theories. Recovering this sense will require us to define a category of models for the theory of Einstein algebras. In order to do this, we need to discuss the "structure-preserving maps" between Einstein algebras.

Let A and B be smooth algebras with $\psi : A \to B$ an algebra homomorphism. Let q be a point in $|B|$ and let $\hat{X}_q \in T_qB$ be an element of the tangent space to B at q. The *pullback* of $\hat{X}_q$ along the homomorphism ψ is the element $\psi^*(\hat{X}_q)$ of $T_{|\psi|(q)}A$ defined by its action $\psi^*(\hat{X}_q)(f) = \hat{X}_q(\psi(f))$ on arbitrary elements $f \in A$. One again uses the correspondence (2.4) between vectors at the point $|\psi|(q)$ in the manifold $G(A)$ and elements of $T_{|\psi|(q)}A$ to verify that indeed $\psi^*(\hat{X}_q) \in T_{|\psi|(q)}A$. The pullback also allows us to use a homomorphism between smooth algebras to transfer other structures between the algebras. In particular, if $\hat{g}$ is a metric on A, the *pushforward* $\psi_*(\hat{g})$ of $\hat{g}$ to B is the map $\hat{g} : \hat{\Gamma}(B) \times \hat{\Gamma}(B) \to B$ defined by

$$\psi_*(\hat{g})(\hat{X}, \hat{Y})(p) = \hat{g}(\psi^*(\hat{X}_p), \psi^*(\hat{Y}_p))$$

for derivations $\hat{X}$ and $\hat{Y}$ on B. We now have the machinery to define the structure-preserving maps between Einstein algebras. If $(A, \hat{g})$ and $(B, \hat{g}')$ are Einstein algebras, an algebra homomorphism $\psi : A \to B$ is an *Einstein algebra homomorphism* if it satisfies $\psi_*(\hat{g}) = \hat{g}'$. Einstein algebra homomorphisms are required to preserve both algebraic structure and the metric structure on the algebras.

We can now define the category of models **EA** for the theory of Einstein algebras. The objects of the category **EA** are Einstein algebras $(A, \hat{g})$, and the arrows are Einstein algebra homomorphisms. Our aim is to prove that **EA** and **GR** are dual categories. We first isolate two facts about the relationship between algebra homomorphisms and smooth maps in the following lemma.

Lemma 2.7. Let $\phi : M \to N$ be a smooth map between manifolds M and N, and $\psi : A \to B$ be an algebra homomorphism between smooth algebras A and B. Then the following both hold:

(1) $\widehat{\varphi_*(X_p)} = \hat{\varphi}^*(\hat{X}_p)$ for every vector X_p at the point $p \in M$;

(2) $\psi^*(\hat{X}_q) = \widehat{|\psi|_*(X_q)}$ for every $\hat{X}_q \in T_q B$.

Proof. Let X_p be a vector at p in the manifold M and $f \in C^\infty(N) = F(N)$. We demonstrate that (1) holds simply by computing the following.

$$\widehat{\varphi_*(X_p)}(f) = \varphi_*(X_p)(f) = X_p(f \circ \phi) = X_p(\hat{\varphi}(f)) = \hat{X}_p(\hat{\varphi}(f)) = \hat{\varphi}^*(\hat{X}_p)(f)$$

The first and fourth equalities follow from the correspondence (2.4), the second equality from the standard geometric definition of the pushforward ϕ_*, the third from the definition of the map $\hat{\phi}$, and the fifth from the algebraic definition of the pullback $\hat{\phi}^*$.

The argument for (2) is perfectly analogous. Let $q \in |B|$ be a point with $\hat{X}_q \in T_q B$ and $f \in A$. We compute the following.

$$\psi^*(\hat{X}_q)(f) = \hat{X}_q(\psi(f)) = X_q(\psi(f)) = X_q(f \circ |\psi|) = |\psi|_*(X_q)(f) = \widehat{|\psi|_*(X_q)}(f)$$

The first equality follows from the algebraic definition of the pullback ψ^*, the second and fifth follow from the correspondence (2.4), the third by the definition of $|\psi|$, and the fourth by the standard geometric definition of the pushforward $|\psi|_*$. □

In conjunction with Theorem 2.5, Lemmas 2.6 and 2.7 allow us to define a pair of translations between the framework of Einstein algebras and the standard framework of general relativity. We first consider the natural way to translate relativistic spacetimes into Einstein algebras. We call this translation J and define it as follows.

- Given a relativistic spacetime (M, g), $J(M, g) = (C^\infty(M), \hat{g})$ is the Einstein algebra with underlying smooth algebra $C^\infty(M)$ and metric $\hat{g}$ defined in Lemma 2.6.
- Given an isometry $\phi : (M, g) \to (M', g')$, $J(\phi)$ is the map $\hat{\phi} : C^\infty(M') \to C^\infty(M)$.

The translation J is perfectly analogous to the contravariant functor F described above. Indeed, as with F we have the following simple result about J.

Lemma 2.8. $J : \mathbf{GR} \to \mathbf{EA}$ is a contravariant functor.

Proof. If (M, g) is an object in **GR**, then it immediately follows that $J(M, g)$ is an object in **EA**. Proposition 2.1 implies that $C^\infty(M)$ is a smooth algebra and Lemma 2.6 implies that $\hat{g}$ is a metric of Lorentz signature on $C^\infty(M)$, so $J(M, g)$ is an Einstein algebra.

Now let $\phi : (M, g) \to (M', g')$ be an isometry. We need to show that $J(\phi) = \hat{\phi} : C^\infty(M') \to C^\infty(M)$ is an Einstein algebra homomorphism. Since ϕ is a smooth map, Lemma 2.3 guarantees that $\hat{\phi} : C^\infty(M') \to C^\infty(M)$ is an algebra homomorphism. It remains to show that

$\hat{\varphi}_*(\hat{g}') = \hat{g}$. Let $\hat{X}$ and $\hat{Y}$ be derivations on $C^\infty(M)$. We compute that

$$\begin{aligned}
\hat{\varphi}_*(\hat{g}')(\hat{X}_p, \hat{Y}_p) &= \hat{g}'(\hat{\varphi}^*(\hat{X}_p), \hat{\varphi}^*(\hat{Y}_p)) \\
&= \hat{g}'(\widehat{\varphi_*(X_p)}, \widehat{\varphi_*(Y_p)}) \\
&= g'(\varphi_*(X_p), \varphi_*(Y_p)) \\
&= \varphi^*(g')(X_p, Y_p) = g(X_p, Y_p) = \hat{g}(\hat{X}_p, \hat{Y}_p)
\end{aligned}$$

for every point $p \in M = |C^\infty(M)|$. The first equality follows from the definition of $\hat{\phi}_*$, the second from Lemma 2.7, the third from the definition of $\hat{g}'$, the fourth from the definition of ϕ^*, the fifth since ϕ is an isometry, and the sixth from the definition of $\hat{g}$. This implies that $\hat{\phi}_*(\hat{g}') = \hat{g}$ and therefore that $J(\phi) = \hat{\phi}$ is an arrow $J(M', g') \to J(M, g)$. One easily verifies that J preserves identities and reverses composition. □

There is also a way to translate from the framework of general relativity into the framework of Einstein algebras. We call this translation K and define it as follows.

- Given an Einstein algebra $(A, \hat{g})$, $K(A, \hat{g}) = (|A|, |\hat{g}|)$ is the relativistic spacetime with underlying manifold $|A| = G(A)$ and metric $|\hat{g}|$ defined in Lemma 2.6.
- Given an Einstein algebra homomorphism $\psi : (A, \hat{g}) \to (A', \hat{g}')$, $K(\psi)$ is the map $|\psi| : |A'| \to |A|$.

The translation K is perfectly analogous to the contravariant functor G described above. And again, we have the following result.

Lemma 2.9. $K : \mathbf{EA} \to \mathbf{GR}$ is a contravariant functor.

Proof. If $(A, \hat{g})$ is an object in **EA**, then it immediately follows that $K(A, \hat{g})$ is an object in **GR**. Indeed, we have already seen that $G(A) = |A|$ is a smooth manifold, and Lemma 2.6 implies that $|\hat{g}|$ is a metric on $|A|$, so $(|A|, |\hat{g}|)$ is a relativistic spacetime.

Now let $\psi : (A, \hat{g}) \to (A', \hat{g}')$ be an Einstein algebra homomorphism. Lemma 2.4 guarantees that $K(\varphi) = |\psi| : |A'| \to |A|$ is a smooth map. It remains to show that $|\psi|^*(|\hat{g}|) = |\hat{g}'|$. By Lemma 2.6, it will suffice to show that $|\psi|^*(g) = g'$. We let X and Y be vector fields on $|A'|$ and compute that

$$\begin{aligned} |\psi|^*(g)(X_p, Y_p) &= g(|\psi|_*(X_p), |\psi|_*(Y_p)) \\ &= \hat{g}(\widehat{|\psi|_*(X_p)}, \widehat{|\psi|_*(Y_p)}) \\ &= \hat{g}(\psi^*(\hat{X}_p), \psi^*(\hat{Y}_p)) \\ &= \psi_*(\hat{g})(\hat{X}_p, \hat{Y}_p) = \hat{g}'(\hat{X}_p, \hat{Y}_p) = g'(X_p, Y_p) \end{aligned}$$

for every point $p \in |A'|$. The first equality follows from the definition of $|\psi|^*$, the second from the definition of $\hat{g}$, the third from Lemma 2.7, the fourth from the definition of ψ_*, the fifth since ψ is an Einstein algebra homomorphism, and the sixth from the definition of $\hat{g}'$. This implies that $|\psi|$ is an isometry and therefore an arrow $K(A', \hat{g}') \to K(A, \hat{g})$. One again easily verifies that K preserves identities and reverses composition. □

We now have the resources necessary to prove our main result. The contravariant functors J and K realize a duality between the category of models for the theory of Einstein algebras and the category of models for general relativity. This result essentially follows as a corollary to Theorem 2.5 along with parts (3) and (4) of Lemma 2.6.

Theorem 2.10. The categories **EA** and **GR** are dual.

Proof. The proof exactly mirrors the proof of Theorem 2.5. We again show that the families of maps $\eta : 1_{\mathbf{EA}} \Rightarrow J \circ K$ and $\theta : 1_{\mathbf{GR}} \Rightarrow K \circ J$ defined in equations (2.1) and (2.2) are natural isomorphisms. It follows from Theorem 2.5 that the naturality squares for η and θ commute, so we need only check that the components of η and θ are isomorphisms.

Let $(A, \hat{g})$ be an object in **EA** and consider the component $\eta_{(A,\hat{g})} : A \to C^\infty(|A|)$. We have already seen in Theorem 2.5 that $\eta_{(A,\hat{g})}$ is an algebra isomorphism. In addition, part (4) of Lemma 2.6 implies that $\eta_{(A,\hat{g})}$ preserves the metric and therefore is an isomorphism between Einstein algebras. A perfectly analogous argument demonstrates that the components $\theta_{(M,g)}$ are isomorphisms between relativistic spacetimes. □

Geroch (1972) defines other structures—analogous to, for instance, tensor fields and covariant derivative operators—in purely algebraic terms, using strategies similar to those used here to define derivations and metrics. With this machinery, he argues, one can express any equation one likes, including Einstein's equation and various matter field equations, in algebraic terms. In this way, one may proceed to do relativity theory using Einstein algebras and structures defined on them, in much the same way that one would using Lorentzian manifolds. The functors J and K, meanwhile, along with the results proved and methods developed here, provide a way of translating between equations relating tensor fields on a Lorentzian manifold (M, g) and the corresponding structures defined on the Einstein algebra $J(M, g)$. Moreover, there is a strong sense in which J and K preserve any possible empirical structure associated with general relativity, on either formulation, since any of the empirical content of general relativity on Lorentzian manifolds will be expressed using invariant geometrical structures such as curves, tensor fields, etc. or their algebraic analogues, and it is precisely this sort of structure that J and K preserve.

2.5 So, is there a spacetime substance or what?

Theorem 2.10 establishes a sense in which the Einstein algebra formalism as it has been defined here has the same structural content as a manifold with metric. This result undercuts Earman's claim that using Einstein algebras lets us eliminate any structure, let alone a that of a spacetime substance, from general relativity. The functor $J : \mathbf{GR} \to \mathbf{EA}$ allows us to "translate" the spacetime manifolds in **GR** into the language of Einstein algebras. So while one may not need to *posit* a spacetime manifold in order to describe a state of the world in terms of Einstein algebras, it is still implicitly there.

As discussed in chapter 1, the TCM verdict depends largely on way in which one chooses to define the categories and functor, but this dependence does not make the results arbitrary. Rather, TCM illuminates the broader connotations of ones prior beliefs about the representational capacities of the formalisms it evaluates. In order to challenge the equivalence verdict, one would need to present a plausible alternative account of the appropriate categories and functors. However, the fact that Einstein algebras were specifically defined to serve all of the same functions as relativistic spacetimes indicates that no viable alternative TCM story will be forthcoming.

An Einstein algebra has all of the building blocks to uniquely define a spacetime manifold. The only difference is that this manifold is not named explicitly in **EA** as it is **GR**. One might want to claim that this is a difference that indeed makes a difference—that what is named in the statement of a theoretical framework has a special ontological status in that framework. To claim this, however, is to give up on the goal of capturing the language independent *structure* of the formalisms that TCM is designed to evaluate.

On a structural conception of the content of formalisms, then, any feature *of the formalism* present in the relativistic spacetime framework is also present for Einstein algebras. This fact can be brought to bear on the relationism-substantivalism debate in a number of ways. If one takes classical general relativity to be committed to the reality of a spacetime, then the theory of Einstein algebras involves a similar commitment. But if one takes the theory of Einstein algebras to in a sense demonstrate the fundamental relational nature of general relativity, then this feature remains when one transitions back from **EA** to **GR** via the functor J. In this way, TCM does not (and cannot) provide a verdict as to whether belief in the truth of general relativity commits one to a belief in the reality of a spacetime substance.

But it does show us that contra Earman, moving to an Einstein algebra framework would not allow us to escape such a commitment.

Chapter 3

Yang-Mills Theory

The main result of this paper was achieved in collaboration with James Weatherall, now published in (Rosenstock and Weatherall, 2016).

3.1 Introduction

Yang-Mills theories generalize the formalism of classical electromagnetism to cover field theories with different symmetry groups, and in their quantized form play a central role in many of our most successful physical theories. Since the introduction of the principal bundle formalism of Yang-Mills theory by Wu and Yang (1975), mathematicians, physicists, and philosophers alike have continued to explore various formalisms and their relative merits. For philosophers of physics, the focus has been on determining which formalism best captures the ontology of the theory, and which features of the formalisms can be understood to represent physically real phenomena.

It has recently been argued by various philosophers, most prominently Richard Healey in his 2007 book *Gauging What's Real*, that formulations of Yang-Mills theories in terms of what are called “holonomies” are preferable to formulations in terms of “principal bundles,” largely on the basis of parsimony considerations. Healey claims that principal bundle formulations posit “surplus structure” relative to holonomy formulations (p. 30), so we should expect that the latter captures the true structure of the world, whereas the former possesses unnecessary mathematical fluff that obscures the physical interpretation. Healey proceeds to describe the meaningful physical differences between a principal bundle and holonomy pictures of the world. For example, the holonomy formalism suggests that properties in Yang-Mills systems are highly non-local, in that they are attributed to curves on spacetime, rather than to spacetime points as they are in the principal bundle picture.

In this chapter, I argue that TCM considerations suggest that contra Healey, the holonomy formalism in a sense possesses at least as much structure as the principal bundle formalism. This indicates that features thought to be readable directly from the principal bundle formalisms, such as “gauge”, are equally structurally present in the holonomy formalism, even when not explicitly named. I begin by describing the principal bundle (section 3.2) and holonomy (3.3) formalisms in the TCM framework. In section 3.4, I establish the functorial relationship between the holonomy and principal bundle formalisms. In section 3.5, I discuss a different formalism—that of Wilson loops—which may add even more structure to the holonomy formalism, but nonetheless points in a promising direction for future research.

3.2 Principal Bundles

A *principal bundle* is a mathematical object $G \to P \xrightarrow{\pi} M$ (abbreviated P) where P is a smooth $(n{+}m)$-dimensional manifold, called the *total space*, M is an n-dimensional connected smooth manifold called the *base space* (here representing space-time), G is an m-dimensional Lie group (the group of gauge symmetries), and π is a projection map from P to M. π is such that at every point x in M, the preimage $\pi^{-1}[x]$ is diffeomorphic to G. This gives us a

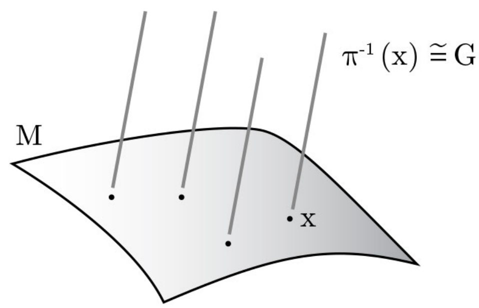


Figure 3.1: A principal bundle.

picture of a principal bundle P as consisting of the manifold M with copies of G associated with each point (see figure 3.1). We require that P is locally a product space of M and G in the sense that every $x \in M$ has a neighborhood $U \subset M$ such that there is a diffeomorphism $\eta : U \times G \to \pi^{-1}[U]$ such that $\pi \circ \eta : (q, g) \mapsto q$ for all $q \in U$ and $g \in G$ (η is called a *local trivialization*). Lastly, there is a smooth right action G on P that preserves the fibers (i.e., for all $u \in P$ and $g \in G$, $\pi(ug) = \pi(u)$) and acts freely and transitively on the fibers, meaning that only the identity element does nothing to the elements of P, and there is an element of G that can take any element of P to any other element of the same fiber.

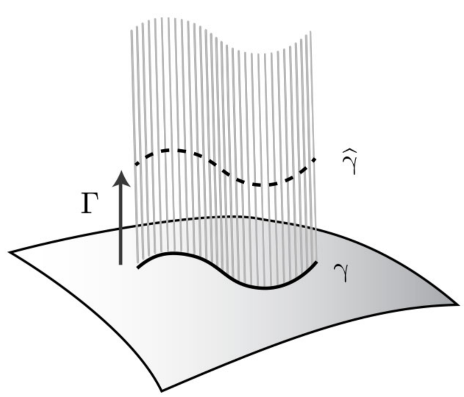


Figure 3.2: A connection.

A *connection* on a principal bundle P is a smooth assignment of a collection of what are called *horizontal lifts* to each space-time curve $\gamma : [0,1] \to M$. Pick a point $\hat{x}$ in the fiber over the initial point $x = \gamma(0)$ of the curve γ. There are in general many ways to associate γ with a "lift" $\hat{\gamma}$ into the fibers over the curve that passes through $\hat{x}$ (i.e., there are many curves $\alpha : [0,1] \to P$ such that $\pi \circ \alpha(t) = \gamma(t)$). A connection picks out precisely one of these for every such curve and every element in the fiber over its initial point (see figure 3.2). Given such a lift, the point $\hat{\gamma}(1)$ is called the *parallel transport* of $\hat{x} = \hat{\gamma}(0)$ along the curve γ according to the connection Γ.

A *principal connection* is a connection Γ that is *equivariant*, i.e., appropriately compatible with the G-action on P in the following sense. Given two elements u and v in a fiber $\pi^{-1}[x]$, where u and v are related via the G-action as $v = ug$ for some $g \in G$, then for every space-time curve γ passing through x, the lifts of γ through u and v are related as $\hat{\gamma}_v(t) = \hat{\gamma}_u(t)g$ for all $t \in [0,1]$. In other words, a principal connection lets us extend the G-action from points in P to lifts of space-time curves.

A principal connection Γ can equivalently be characterized by a one-form ω on P that takes values in $\mathfrak{g}$, the Lie algebra of G. On this characterization, a connection picks out a preferred decomposition of each tangent vector $\xi \in T_uP$ at each point $u \in P$ into the part that's "parallel to M" (the *horizontal* component) and the part that's "parallel to the fiber" (the *vertical* component). ω thus tells us which vectors in $T_{\pi(u)}M$ correspond "lifts" of vectors in T_uM, which lets us characterize horizontal lifts of curves in terms of infinitesimal directions of the curves at a point.

A *principal bundle isomorphism* from $G \to P \xrightarrow{\pi} M$ to $G' \to P' \xrightarrow{\pi'} M'$ is a pair (f,g) consisting of a diffeomorphism $f : P \to P'$, and a Lie group isomorphism $g : G \to G'$ such that $f(xa) = f(x)g(a)$. (f,g) *preserves the connection* ω on P if the connection ω' on P' is such that $f^*\omega' = g_* \circ \omega$, where $\tilde{g} : \mathfrak{g} \to \mathfrak{g}'$ is the Lie algebra isomorphism induced by g.

The connection ω, and its *curvature form* $D\omega$ (where D is the exterior covariant derivative), can also be represented as differential forms on M, rather than P. However, this requires pulling back along a *(local) section*, or map $\sigma : U \to P$ where $U \subseteq M$ and $\pi \circ \sigma = \mathrm{id}_U$. There will not in general be a section definable on all of M. Given a choice of local section σ, the pullback $\sigma^*\omega : U \to \mathfrak{g}$ of the connection along the section is a *Yang-Mills potential*, and $\sigma^*(D\omega)$ a *Yang Mills field*.

Both the Yang-Mills potential and field will in general depend on the choice of section used to represent them, and this choice is referred to as a *choice of gauge*. When it is said that the physically meaningful content of Yang-Mills theory must be gauge-independent, it is thus meant that observables of the theory cannot depend on a particular choice of section used to represent ω on M. A *gauge transformation* is a change in choice of gauge, i.e. a smooth map $t : P \to P$ such that $\pi \circ t(x) = \pi(x)$ for all $x \in P$, which takes a section σ to a section $t \circ \sigma$. Equivalently, one can think of a gauge transformation as a principal bundle automorphism while holding the section fixed.

The *category of principal bundle models* of Yang-Mills theory is the category **PC** whose objects are principal bundles with connection and whose arrows are principal bundle isomorphisms that preserve the connection. Given the formalism and its intended application, there is not much room for disagreement on how this category is defined.

It is sometimes suggested that rather than allowing general principal bundle isomorphisms, only *vertical isomorphisms*, or isomorphisms that preserve the fibers and act as the identity on M, should be admitted. This might be motivated by substantivalism about spacetime—the idea that once we pick a representation of space-time as a manifold M, any automorphism of that manifold changes it, since it changes which properties are assigned to which points. However, the formalism of classical field theory does not differentiate between space-time manifolds related by diffeomorphisms, so long as the relevant fields are pushed forward along the diffeomorphism. To disallow such isomorphisms would be to misunderstand the formalism and its role in representing field theories (see Weatherall (2016b) for more on this line).

Another possible contention might be that principal bundle models of Yang Mills theory should also include a privileged choice of (local) section, and further require that arrows preserve this choice of section. Call this category $\mathbf{PC}^*$. Let $F : \mathbf{PC}^* \to \mathbf{PC}$ be the functor that "forgets" the favorite section for each object in $\mathbf{PC}^*$, and maps arrows accordingly. This map is clearly surjective on objects. Since F acts as the identity on arrows, it is also faithful. Let (P, ω, σ) be an object in $\mathbf{PC}^*$. Then $F(P, \omega, \sigma) = (P, \omega)$, which has automorphisms that do not preserve σ, since such a transformation would take you to a distinct object in $\mathbf{PC}^*$. Thus F is not full, so F forgets structure—namely, the structure of a preferred section.

3.3 Holonomies

I will first define the notion of a holonomy in the context of the principal bundle formalism, and then demonstrate how the notion can be formulated independently of a principal bundle or connection.

One perhaps surprising aspect of parallel transport is its path-dependence: two space-time curves with the same endpoints that are lifted to the same initial point in the principal bundle might then have different endpoints in the fiber above their shared final point. If parallel transport is path-independent, at least locally, the connection is said to be *flat*.

The notion of holonomy makes precise the sense in which the connection exhibits path dependence. Given a principal bundle P with connection Γ, a closed curve $\gamma : [0,1] \to M$ starting and ending at $x \in M$, and a point $u \in \pi^{-1}[x]$, the *holonomy* of γ (relative to P, Γ, and u) is the element g of G that relates the initial point u of the lift $\hat{\gamma}_u$ of γ to its final point, i.e., $\hat{\gamma}(1) = ug = \hat{\gamma}(0)g$ (see figure 3.3). The holonomy g of γ thus characterizes the sense in which the space traversed by γ is "curved," or fails to be flat.

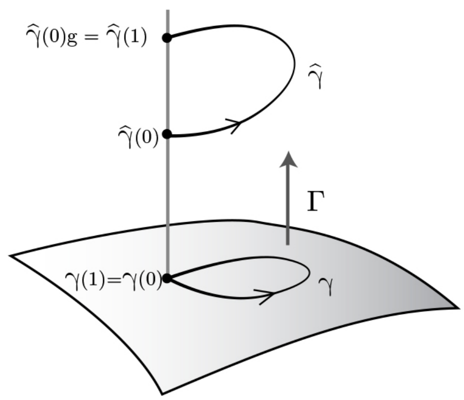


Figure 3.3: A holonomy assignment.

In general the holonomy g of γ will depend on one's choice of lift point $u \in \pi^{-1}[x]$. However, if we look at the *holonomy map* $H_{\Gamma,u} : L_x \to G$ from closed curves based at $x = \pi(u)$ to G given by the holonomies of curves relative to Γ and u, the resulting assignments will be related to those induced by another element $u' \in \pi^{-1}[x]$ by a Lie group isomorphism (specifically, an inner automorphism). Thus we can say that the holonomy map gives us the same information about closed curves regardless of lift point, insofar as isomorphisms preserve all relevant information.

There is also a sense in which it does not matter what base point x we choose. Consider a closed curve γ based at a distinct point $x' \in M$. If we fix a choice of curve α from x to x', then the (reparameterized) composition of curves $\alpha^{-1} \bullet \gamma \bullet \alpha$ is in L_x, and $H_{\Gamma,\hat{\alpha}_u(1)}(\gamma) = H_{\Gamma,u}(\alpha^{-1} \bullet \gamma \bullet \alpha)$. This identification will depend on one's choice of connecting curve α, but again, all such assignments will be related to one another by an inner automorphism.

A holonomy map can be defined independently of an underlying principal bundle and connection as follows. Pick a point $x \in M$. We'll say that two curves $\gamma_1, \gamma_2 \in L_x$ are *thinly equivalent*, written $\gamma_1 \sim \gamma_2$, if there exists a homotopy h of $\gamma_1^{-1} \bullet \gamma_2$ to the null curve

$\mathrm{id}_x : [0,1] \mapsto x$ such that the image of h is contained in the image of $\gamma_1^{-1} \bullet \gamma_2$. In other words, curves are thinly equivalent whenever their images differ by at most curves of empty interior. Note that thinly equivalent curves will always have the same holonomies by the equivariance of principal connections.

We can now define a *generalized holonomy map* on M with reference point x and structure group G to be a map $H : L_x \to G$ such that (1) for any $\gamma_1, \gamma_2 \in L_x$, if $\gamma_1 \sim \gamma_2$, then $H(\gamma_1) = H(\gamma_2)$; (2) for any $\gamma_1, \gamma_2 \in L_x$, $H(\gamma_1 \bullet \gamma_2) = H(\gamma_1)H(\gamma_2)$; and (3) H is smooth in the appropriate sense (See Barrett (1991) and Caetano and Picken (1994), who each present slightly different smoothness conditions.). We will thus call a *holonomy model* of a Yang-Mills theory with structure group G a triple (M, x, H) consisting of the space-time manifold M, base-point $x \in M$, and generalized holonomy map $H : L_x \to G$.

As is apparent from the way in which holonomy maps were defined from principal bundles with connections, there will be a holonomy map corresponding to every principal bundle with a connection. Moreover, Barrett (1991) showed how a principal bundle with connection can be reconstructed from a generalized holonomy map (as did (Caetano and Picken, 1994) for a different definition of generalized holonomy map). These two processes are inverse, giving a bijective correspondence between principal bundle and holonomy models of Yang-Mills theory. Nonetheless, a few philosophers of physics—most notably Richard Healey in his 2007 book *Gauging What's Real*—have argued that holonomy models have "less structure" than and are thus preferable to principal bundle models. The argument rests on the claim that holonomy models lack the excess "structure" that principal bundle models have in the form of choice of gauge, which is not physically meaningful.

This is where TCM comes in. Theorem 3.2 sharpens the results of Barrett (1991) and Caetano and Picken (1994) by showing that the relationship between the sets of models of the two interpretations is more than just a bijection, it is, for a definition of the category **Hol** of holonomy models, a (definable) categorical equivalence. Since this gives a precise and well-motivated sense in which principal bundle models do not have excess structure relative to holonomy models, those who still want to claim otherwise must undercut either the fact that an equivalence of categories is the appropriate notion of structural comparison, or that this **Hol** is the appropriate category.

The right choice of "holonomy isomorphism" to use for defining the category **Hol** of holonomy isomorphisms is a subtler business than in **PC**. We want it to account for the three basic ways in which two holonomy models (M, x, H) and (M', x', H') can be said to represent the "same" holonomy data: (1) M and M' can be related by a diffeomorphism—since diffeomorphism is the notion of isomorphism we use for manifolds, this shouldn't change the information encoded in M; (2) if $M' = M$, H' might be the "translation" of the holonomies at x to x' via some curve α; and (3) the values taken by H and H' can be related by a Lie group isomorphism—this can be taken to correspond to a change in "lift point" in the principal bundle language, or merely a change in how we're using G to represent holonomy data.

We define holonomy isomorphism for their category **Hol** as follows. Let $H : L_x \to G$ and $H' : L_{x'} \to G$ be (generalized) holonomy maps on manifolds M and M'. A *holonomy isomorphism* from H to H' is an ordered triple $(\Psi, \underline{\alpha}, \phi)$ where $\Psi : M \to M'$ is a diffeomorphism, $\phi : G \to G$ is a Lie group isomorphism, and $\underline{\alpha}$ is an equivalence class of piece-wise smooth curves $\alpha : [0, 1] \to M$ satisfying $\alpha(0) = \Psi^{-1}(x')$ and $\alpha(1) = x$, which are all such that for any $\gamma \in L_x$, $\phi \circ H(\gamma) = H'(\Psi \circ (\alpha^{-1} \bullet \gamma \bullet \alpha))$.

3.4 A PSS Analysis

In this section, I prove that the functor $C_B : \mathbf{Hol} \to \mathbf{PC}$, given by the Barrett-Caetano-Picken reconstruction, forgets *structure*. This functorial relationship indicates that the principal bundle formalism does not include any additional structure absent from the holonomy formulation, for the general reasons suggested in chapter 1. This result implies that there is no "gauge structure" present in the principal bundle formalism which can be eliminated by instead using the holonomy formalism. This undercuts Healey's primary argument for the relative parsimony of the holonomy formalism,

This functor does split. The functor D_u : **PC** → **Hol**, that takes principle bundles with connection to holonomy maps defined at the point u in that bundle, is inverse to C_B and forgets *stuff*. The stuff that D_u forgets can be thought of as the rest of the principle bundle, beyond the holonomy sub-bundle. For some principle bundles, the holonomy sub-bundle is the entire bundle, so no stuff is forgotten for these objects. So there is a different sense in which the holonomy can be thought of as more parsimonious than the principle bundle formalism, in that it has "less stuff". However, this is easily ameliorated by considering only the holonomy sub-bundles rather than entire bundles in the principal bundle formalism. So it is not principal bundles per se that fail to be parsimonious, only the fact that the principal bundles used are sometimes larger than strictly necessary.

The structure that C_B forgets is indicated by the extra information that is required to define an inverse D_u—namely, a point $u \in P$. In this sense, moving from **PC** to **Hol** requires picking out a preferred point, whereas in **PC** no point has special status relative to any other.

3.4.1 Proofs

Theorem 3.1. The Barrett reconstruction functor C : **Hol** → **PC** forgets only structure.

The proof of Theorem 3.1 depends on the following result concerning the notion of holonomy isomorphism we will presently define. This result is of some interest in its own right.

Theorem 3.2. Let $G \to P \overset{\pi}{\to} M$ and $G' \to P' \overset{\pi'}{\to} M'$ be principal bundles with principal connections Γ and Γ' respectively, and suppose that M and M' are connected. Suppose there are points $u \in P$ and $u' \in P'$ such that the holonomy maps based at u and u' are isomorphic. Then there is a connection-preserving principal bundle isomorphism between P and P'.

Our proof of Theorem 3.2 will depend on the following three lemmas. In what follows $T_{\Gamma,\gamma}(u)$ denotes the parallel transport via a connection Γ on a principal bundle P of a point u along a curve $\gamma : [0,1] \to M$ which is such that $\gamma(0) = \pi(u)$. In other words, $T_{\Gamma,\gamma}(u) = \hat{\gamma}_u(1)$.

Lemma 3.3. Let $G \to P \overset{\pi}{\to} M$ be a principal bundle and let Γ be a principal connection on it. Then for all $x \in M$, $u \in \pi^{-1}[x]$, $\gamma \in L_x$, $g \in G$, and all piece-wise smooth curves $\alpha, \alpha' : [0,1] \to M$ such that $\alpha(0) = \alpha'(0) = x$ and $\alpha(1) = \alpha'(1)$, the following hold:

(a) $T_{\Gamma,\alpha^{-1}\bullet\alpha'}(u) = T_{\Gamma,\alpha^{-1}}(T_{\Gamma,\alpha'}(u))$, where α^{-1} is the reverse orientation of α.

(b) $T_{\Gamma,\alpha^{-1}}(T_{\Gamma,\alpha'}(u)) = u$ iff $T_{\Gamma,\alpha}(u) = T_{\Gamma,\alpha'}(u)$

(c) $H_{\Gamma,u}(\gamma) = e_G$, the identity element of G, iff $T_{\Gamma,\gamma}(u) = u$

(d) $T_{\Gamma,\alpha}(ug) = T_{\Gamma,\alpha}(u)g$

Proof. (a) and (b) follow from the fact that every curve α has a unique horizontal lift $\hat{\alpha}_u$ which is such that $\hat{\alpha}_u(0) = u$. (c) follows from the definition of holonomy map. (d) follows from the equivariance of the connection under the right action of G on P. □

Lemma 3.4. Let $G \to P \overset{\pi}{\to} M$ be a principal bundle and let Γ a principal connection on it. Let $\alpha : [0,1] \to M$ be a piece-wise smooth curve such that $\alpha(0) = x$ and $\alpha(1) = x'$. Then for all $u \in \pi^{-1}[x']$ and all $\gamma \in L_{x'}$, if $v = T_{\Gamma,\alpha^{-1}}(u) \in \pi^{-1}[x]$, then

$$H_{\Gamma,u}(\gamma) = H_{\Gamma,v}(\alpha^{-1} \bullet \gamma \bullet \alpha)$$

Proof. Suppose $H_{\Gamma,u}(\gamma) = g \in G$, i.e. that $T_{\Gamma,\gamma}(u) = ug$. Then by Lemma 3.3 (b) and (d), $vg = T_{\Gamma,\alpha^{-1}}(u)g = T_{\Gamma,\alpha^{-1}}(ug) = T_{\Gamma,\alpha^{-1}}T_{\Gamma,\gamma}(u) = T_{\Gamma,\alpha^{-1}}T_{\Gamma,\gamma}T_{\Gamma,\alpha}(v) = T_{\Gamma,\alpha^{-1}\bullet\gamma\bullet\alpha}(v)$. Therefore $H_{\Gamma,v}(\alpha^{-1} \bullet \gamma \bullet \alpha) = g$ □

In the following lemma, we make use of the *holonomy sub-bundle* $\Phi_{\Gamma,u} \to P_{\Gamma,u} \xrightarrow{\tilde{\pi}} M$ associated with a point $u \in P$ and principal connection Γ on a principal bundle $G \to P \xrightarrow{\pi} M$, as discussed in detail §II.7 of Kobayashi and Nomizu (1996). This is the bundle consisting of all points of P that may be joined to $u \in P$ by a horizontal curve. The Reduction Theorem (Theorem II.7.1 of Kobayashi and Nomizu (1996)) establishes the following about this bundle:

1. $\Phi_{\Gamma,u} \to P_{\Gamma,u} \xrightarrow{\tilde{\pi}} M$ is a reduced sub-bundle of $G \to P \xrightarrow{\pi} M$ with the holonomy group $\Phi_{\Gamma,u}$ as its structure group and with $\tilde{\pi} = \pi_{\restriction P_{\Gamma,u}}$ (and similarly $P'_{\Gamma',u'}$ is a reduction of P').

2. The connection Γ is reducible to a connection $\tilde{\Gamma} = \Gamma_{\restriction\tilde{\pi}}$ on $P_{\Gamma,u}$ (and similarly, Γ' reduces to $\tilde{\Gamma}' = \Gamma'_{\restriction\tilde{\pi}'}$).

That $P_{\Gamma,u}$ is a reduced bundle of P means in particular that $\Phi_{\Gamma,u}$ is a Lie subgroup of G and that each element of P may be written (not necessarily uniquely) as xa for some $x \in P_{\Gamma,u}$ and $a \in G$.

Lemma 3.5. Let $G \to P \xrightarrow{\pi} M$ and $G' \to P' \xrightarrow{\pi'} M'$ be principal bundles with principal connections Γ and Γ' respectively, with M and M' connected. Let $\Phi_{\Gamma,u} \to P_{\Gamma,u} \xrightarrow{\tilde{\pi}} M$ and $\Phi'_{\Gamma',u'} \to P'_{\Gamma',u'} \xrightarrow{\tilde{\pi}'} M'$ be the holonomy sub-bundles of P and P' at u and u', respectively, and $\tilde{\Gamma}$ and $\tilde{\Gamma}'$ be the restrictions of Γ and Γ' to $P_{\Gamma,u}$ and $P'_{\Gamma',u'}$, respectively. If there is a principal bundle isomorphism $(f, \Psi, \phi_{\restriction\Phi_{\Gamma,u}}) : P_{\Gamma,u} \to P'_{\Gamma',u'}$ that preserves the connections $\tilde{\Gamma}$ and $\tilde{\Gamma}'$, where $\Psi : M \to M'$ is a diffeomorphism and $\phi : G \to G'$ is a Lie group isomorphism, then $(f, \Psi, \phi_{\restriction \Phi_{\Gamma,u}})$ can be extended to a principal bundle isomorphism $(F, \Psi, \phi) : P \to P'$ that preserves Γ and Γ'.

Proof. Define $F : P \to P'$ from f as:

$$F(pg) := f(p)\phi(g) \text{ for } p \in P_{\Gamma,u},\ g \in G$$

To prove that (F, Ψ, ϕ) is a principal bundle isomorphism, we must show that F is well-defined and a diffeomorphism, and that the following identities hold:

1. $\pi' \circ F = \Psi \circ \pi$
2. $\pi \circ F^{-1} = \Psi^{-1} \circ \pi'$
3. For all $v \in P$, $g \in G$, $F(vg) = F(v)\phi(g)$

Finally, we must show that (F, Ψ, ϕ) preserves Γ. We do this by showing that the bundles agree, via the transformation (F, Ψ, ϕ), on which curves are horizontal.

To see that F is well-defined, consider any $v \in P$, and suppose there are $x, y \in P_{\Gamma,u}$ and $g, h \in G$ such that $v = xg = yh$. Then $x = yhg^{-1}$, and hence

$$F(xg) = F((yhg^{-1})(g)) = f(yhg^{-1})\phi(g) = f(y)\phi(h)\phi(g^{-1})\phi(g) = f(y)\phi(h) = F(yh)$$

To show that F is also a diffeomorphism, it is sufficient to show that F is bijective and that it is locally a diffeomorphism. First suppose $F(v) = F(w)$ for some $v, w \in P$. Then by the definition of F, $\pi(v) = \pi(w)$, so we may write $v = xg$ and $w = xh$ for the same $x \in P_{\Gamma,u}$. Thus $f(x)\phi(g) = F(v) = F(w) = f(x)\phi(h)$, but since ϕ is an isomorphism, this implies that $g = h$ and hence $v = xg = yh = w$. Thus F is injective. Now consider any $v' \in P'$. Write

$v' = x'g'$ for some $x' \in P'_{\Gamma',u'}$, $g' \in G'$. Then $F(f^{-1}(x')\phi^{-1}(g')) = x'g' = v'$. Since f and ϕ are bijections, $f^{-1}(x')\phi^{-1}(g')$ is a well-defined element of P. So F is bijective.

Finally, let $v \in P$, and let $U \subset M$ be a neighborhood of $\pi(v)$ which is such that a local trivialization of π is defined on U and a local trivialization of π' is defined on $\Psi[U]$. Then there is a local section $\sigma : U \to P_{\Gamma,u}$, and $f \circ \sigma \circ \Psi^{-1}$ is a local section of $P'_{\Gamma',u'}$ on $\Psi[U]$. Then for $p \in \pi^{-1}[U]$,

$$F(p) = F(\sigma \circ \pi(p)\theta(p)) = f \circ \sigma \circ \pi(p)\phi \circ \theta(p),$$

where $\theta : \pi^{-1}[U] \to G$ as $p \mapsto a$, where a is the unique element of G such that $p = \sigma(\pi(p))a$. To see that θ is smooth, let $\xi : \pi^{-1}[U] \to U \times G$ be a local trivialization of P. Then

$$\theta(p) = ((\mathrm{proj}_R \circ \xi \circ \sigma \circ \pi)(p))^{-1}(\mathrm{proj}_R \circ \xi)(p)$$

where $\mathrm{proj}_R : U \times G \to G$ acts as $(z, b) \mapsto b$. Thus $F_{\restriction \pi^{-1}[U]}$ is the product of compositions of smooth maps, and is hence smooth. The argument for its inverse follows by analogy, once one notes that $F^{-1}(x'g') = f^{-1}(x')\phi^{-1}(g')$. This completes the argument that F is a diffeomorphism.

We now confirm that the identities 1-3 above hold. Let $v \in P$. Then $v = xg$ for some $x \in P_{\Gamma,u}$ and $g \in G$. Since f is an isomorphism and $\pi(v) = \pi(x)$,

$$\pi' \circ F(v) = \pi'(f(x)\phi(g)) = \pi'(f(x)) = \Psi(\pi(x)) = \Psi(\pi(v)).$$

So $\pi' \circ F = \Psi \circ \pi$. An identical argument establishes that $\pi \circ F^{-1} = \Psi^{-1} \circ \pi'$. Now suppose we have some $v \in P$ and $g \in G$. Then $v = xh$ for some $x \in P_{\Gamma,u}$ and $h \in G$. It follows that

$$F(vg) = F(xhg) = f(x)\phi(hg) = f(x)\phi(h)\phi(g) = F(v)\phi(g).$$

So $F(vg) = F(v)\phi(g)$, and thus (F, Ψ, ϕ) is a principal bundle isomorphism.

It remains to show that (F, Ψ, ϕ) preserves Γ. Let γ be a smooth curve in M, $v \in \pi^{-1}(\gamma(0))$, and suppose $v = xg$, $x \in P_{\Gamma,u}$, $g \in G$. Since Γ is a principal connection, the lifts of γ to x and v are related as $\hat{\gamma}_v(t) = \hat{\gamma}_x(t)g$. Since f takes $\tilde{\Gamma}$ to $\tilde{\Gamma}'$, we have that

$$F(\hat{\gamma}_v(t)) = F(\hat{\gamma}_x(t)g) = f(\hat{\gamma}_x(t))\phi(g) = \widehat{\Psi \circ \gamma}_{f(x)}(t)\phi(g) = \widehat{\Psi \circ \gamma}_{F(v)}(t).$$

Thus Γ and Γ' agree on horizontal curves. □

We now turn to the principal result of this section, which we restate here for convenience.

Theorem 3.1. Let $G \to P \xrightarrow{\pi} M$ and $G' \to P' \xrightarrow{\pi'} M'$ be principal bundles with principal connections Γ and Γ' respectively, and suppose that M and M' are connected. Suppose there are points $u \in P$ and $u' \in P'$ such that the induced holonomy maps based at u and u' are isomorphic. Then there is a connection-preserving principal bundle isomorphism between P and P'.

Proof. We first show that there is a principal bundle isomorphism $(f, \Psi, \phi) : P_{\Gamma,u} \to P'_{\Gamma',u'}$ that preserves $\tilde{\Gamma}$, where $\Phi_{\Gamma,u} \to P_{\Gamma,u} \xrightarrow{\tilde{\pi}} M$ and $\Phi'_{\Gamma',u'} \to P'_{\Gamma',u'} \xrightarrow{\tilde{\pi}'} M'$ are the holonomy sub-bundles of P and P' at u and u', respectively, and $\tilde{\Gamma}$ and $\tilde{\Gamma}'$ are the restrictions of Γ and Γ' and $P_{\Gamma,u}$ and $P'_{\Gamma',u'}$, respectively. We then invoke Lemma 3.5 to extend (f, Ψ, ϕ) to a principal bundle isomorphism $(F, \Psi, \phi) : P \to P'$ that preserves Γ.

First, since $H_{\Gamma,u}$ and $H'_{\Gamma',u'}$, the holonomy maps induced by Γ and Γ' and based at u and u', respectively, are isomorphic by assumption, there must be some holonomy isomorphism $(\Psi, \underline{\alpha}, \phi) : H_{\Gamma,u} \to H'_{\Gamma',u'}$. Let $z := T_{\Gamma,\alpha^{-1}}(u) \in \pi^{-1}(\alpha(0))$, where $\alpha \in \underline{\alpha}$. (Note that $z \in P_{\Gamma,u}$, and moreover $P_{\Gamma,u} = P_{\Gamma,z}$, i.e., every element of $P_{\Gamma,u}$ can be connected to z via some piece-wise smooth, horizontal curve). Define $f : P_{\Gamma,u} \to P'_{\Gamma',u'}$ as follows:

(i) $f(z) := u'$

(ii) For any $v \in P_{\Gamma,u}$, pick some piece-wise smooth curve $\beta_v \in C_{M,\pi(z)}$ (where $C_{M,\pi(z)}$ denotes the set of piece-wise smooth space-time curves $\gamma : [0,1] \to M$ such that $\gamma(0) = \pi(z) = \alpha(0)$) such that $v = T_{\tilde{\Gamma},\beta_v}(z)$, the parallel transport in $P_{\Gamma,u}$ of z along β_v according to the connection $\tilde{\Gamma}$. Then set $f(v) := T'_{\tilde{\Gamma}',\Psi\circ\beta_v}(u')$, the parallel transport in $\tilde{\pi}$ of u' along $\Psi \circ \beta_v$.

We claim that the triple (f, Ψ, ϕ) realizes the desired principal bundle isomorphism. To prove this, we must show that f is well-defined, a diffeomorphism, and that the following identities hold:

1. $\tilde{\pi}' \circ f = \Psi \circ \tilde{\pi}$
2. $\tilde{\pi} \circ f^{-1} = \Psi^{-1} \circ \tilde{\pi}'$
3. For all $v \in P_{\Gamma,u}$, $g \in \Phi_{\Gamma,u}$, $f(vg) = f(v)\phi(g)$

Finally, we must show that (f, Ψ, ϕ) preserves the reduced connection $\tilde{\Gamma}$.

We begin by showing that f is well-defined. Consider any point $v \in P_{\Gamma,u}$. Suppose the curves β and $\beta' \in C_{M,\pi(z)}$ are such that $T_{\tilde{\Gamma},\beta}(z) = T_{\tilde{\Gamma},\beta'}(z) = v$. We want to show that $T'_{\tilde{\Gamma}',\Psi\circ\beta}(u') = T'_{\tilde{\Gamma}',\Psi\circ\beta'}(u')$. Let β^{-1} denote the reverse orientation of β, and e_G the identity element of G (and hence of $\Phi_{\Gamma,z}$ and $\Phi_{\Gamma,u}$). By Lemma 3.3 (a) and (b), $T_{\tilde{\Gamma},\beta^{-1}\bullet\beta'}(z) = T_{\tilde{\Gamma},\beta^{-1}}(T_{\tilde{\Gamma},\beta'}(z)) = T_{\tilde{\Gamma},\beta^{-1}}(v) = z$. Thus by Lemma 3.3 (c), $H_{\Gamma,z}(\beta^{-1} \bullet \beta') = e_G$. Since ϕ is a Lie group isomorphism, we also know that $\phi(e_G) = e_{G'}$. By Lemma 3.4, then, we know that $e_G = H_{\Gamma,z}(\beta^{-1} \bullet \beta') = H_{\Gamma,u}(\alpha \bullet \beta^{-1} \bullet \beta' \bullet \alpha^{-1}) = H_{\Gamma,u}(\bar{\alpha}^{-1}(\beta^{-1} \bullet \beta'))$, where $\bar{\alpha}$ is as in the definition of holonomy isomorphism in section 3.3. Since $(\Psi, \underline{\alpha}, \phi)$ is a holonomy isomorphism, we know that $e_{G'} = \phi \circ H_{\Gamma,u}(\bar{\alpha}^{-1}(\beta^{-1} \bullet \beta')) = (H_{\Gamma',u'} \circ \psi \circ \bar{\alpha})(\bar{\alpha}^{-1}(\beta^{-1} \bullet \beta')) = H_{\Gamma',u'}(\Psi \circ (\beta^{-1} \bullet \beta'))$. This tells us that $u' = T'_{\tilde{\Gamma}',\Psi\circ(\beta^{-1}\bullet\beta')}(u') = T'_{\tilde{\Gamma}',\Psi\circ\beta^{-1}}(T'_{\tilde{\Gamma}',\Psi\circ\beta'}(u'))$. By Lemma 3.3 (b), this implies that $T'_{\tilde{\Gamma}',\Psi\circ\beta}(u') = T'_{\tilde{\Gamma}',\Psi\circ\beta'}(u')$. So f is well-defined.

We now show that f is bijective. (Later we will also show that f and f' are smooth, completing the proof that f is a diffeomorphism.) Let $v, w \in P_{\Gamma,u}$, and suppose $f(v) = f(w)$. We want to show that $v = w$. Since $f(v) = f(w)$, we know that $T'_{\tilde{\Gamma}',\Psi\circ\beta_v}(u') = f(v) = f(w) = T'_{\tilde{\Gamma}',\Psi\circ\beta_w}(u')$. By Lemma 3.3 (a) and (b) and the fact that Ψ is a diffeomorphism, we get that $u' = T'_{\tilde{\Gamma}',(\Psi\circ\beta_v)^{-1}}(T'_{\tilde{\Gamma}',\Psi\circ\beta_w}(u')) = T'_{\tilde{\Gamma}',(\Psi\circ\beta_v)^{-1}\bullet(\Psi\circ\beta_w)}(u') = T'_{\tilde{\Gamma}',\Psi\circ(\beta_v^{-1}\bullet\beta_w)}(u')$. Thus by Lemma 3.3 (c) we get that $H_{\Gamma',u'}(\Psi \circ (\beta_v^{-1} \bullet \beta_w)) = e_{G'}$. Since $(\Psi, \underline{\alpha}, \phi)$ is a holonomy isomorphism, this implies that $\phi(H_{\Gamma,u}(\bar{\alpha}^{-1}(\beta_v^{-1} \bullet \beta_w))) = e_{G'}$, which, since ϕ is a Lie group isomorphism, implies that $H_{\Gamma,u}(\bar{\alpha}^{-1}(\beta_v^{-1} \bullet \beta_w)) = e_G$. By Lemma 3.4, then, $H_{\Gamma,z}(\beta_v^{-1} \bullet \beta_w) = e_G$. Thus by Lemma 3.3 (c), $v = T_{\tilde{\Gamma},\beta_v}(z) = T_{\tilde{\Gamma},\beta_w}(z) = w$. So f is injective. Now let $w' \in P'_{\Gamma',u'}$, and let the curve $\beta' \in C_{M',\pi'(u')}$ be such that $T'_{\tilde{\Gamma}',\beta'}(u') = w'$. Then there is a unique $v \in P_{\Gamma,u}$ such that $v = T_{\tilde{\Gamma},\Psi^{-1}\circ\beta'}(z)$. Then $f(v) = T'_{\tilde{\Gamma}',\Psi\circ\alpha_v}(u') = T'_{\tilde{\Gamma}',\Psi\circ(\Psi^{-1}\circ\beta')}(u') = T'_{\tilde{\Gamma}',\beta'}(u') = w'$. (The second equality follows from fact that f is well-defined.) It follows that f is bijective.

We will now establish identities 1-3. Let $v \in P_{\Gamma,u}$. Then

$$\tilde{\pi}'(f(v)) = \tilde{\pi}'(T'_{\tilde{\Gamma}',\Psi\circ\beta_v}(u')) = (\Psi \circ \beta_v)(1) = \Psi(\beta_v(1)) = \Psi(\tilde{\pi}(v)).$$

So $\tilde{\pi}' \circ f = \Psi \circ \tilde{\pi}$. By identical reasoning, $\tilde{\pi} \circ f^{-1} = \Psi^{-1} \circ \tilde{\pi}'$. Finally, let $v \in P_{\Gamma,u}$ and $g \in \Phi_{\Gamma,u}$. First note that by Lemma 3.3 (d) and the well-definedness of f, we can assume without loss of generality that $\beta_{vg} = \beta_v \bullet \beta_{zg}$. By Lemma 3.3 (a), $f(vg) = T'_{\tilde{\Gamma}',\Psi\circ(\beta_v\bullet\beta_{zg})}(u') = T'_{\tilde{\Gamma}',\Psi\circ\beta_v}(T'_{\tilde{\Gamma}',\Psi\circ\beta_{zg}}(u'))$. By the definition of holonomy isomorphism, $T_{\tilde{\Gamma}',\Psi\circ\beta_{zg}}(u') = u'H_{\Gamma',u'}(\Psi\circ\beta_{zg}) = u'H_{\Gamma',u'}(\Psi \circ \bar{\alpha} \circ (\alpha \bullet \beta_{zg} \bullet \alpha^{-1})) = u'\phi(H_{\Gamma,u}(\alpha \bullet \beta_{zg} \bullet \alpha^{-1})) = u'\phi(H_{\Gamma,z}(\beta_{zg})) = u'\phi(g)$. Plugging this equality into the last one, and using Lemma 3.3 (d), we get: $f(vg) = T'_{\tilde{\Gamma}',\Psi\circ\beta_v}(u'\phi(g)) = T'_{\tilde{\Gamma}',\Psi\circ\beta_v}(u')\phi(g) = f(v)\phi(g)$.

Next we show that f preserves Γ. It suffices to show that for all piece-wise smooth curves $\gamma : [0, 1] \to M$ and all $w \in \pi^{-1}(\gamma(0))$, $f(T_{\tilde{\Gamma},\gamma}(w)) = T'_{\tilde{\Gamma},\Psi\circ\gamma}f(w)$. But this follows easily from the definition of f: $f(T_{\tilde{\Gamma},\gamma}(w)) = T'_{\tilde{\Gamma},\Psi\circ\beta_{T_{\tilde{\Gamma},\gamma}(w)}}(u') = T'_{\tilde{\Gamma},\Psi\circ(\gamma\bullet\beta_w)}(u') = T'_{\tilde{\Gamma},\Psi\circ\gamma}(T'_{\tilde{\Gamma},\Psi\circ\beta_w}(u')) = T'_{\tilde{\Gamma},\Psi\circ\gamma}(f(w))$.

To complete the proof, we have only to show that f and f^{-1} are smooth. Then f will be a diffeomorphism, and (f, Ψ, ϕ) will be a principal bundle isomorphism that preserves Γ. Let $v \in P_{\Gamma,u}$ and let $V \subseteq M$ an open neighborhood of $x = \tilde{\pi}(v)$ on which a local trivialization of $P_{\Gamma,u}$ is defined. Let V' be a neighborhood of $\Psi(x)$ on which a local trivialization of $P'_{\Gamma',u'}$ is defined. Let g be a metric on M, $g' = \Psi_*(g)$. Let U be an open subset of $V \cap \Psi^{-1}[V']$ (containing x) on which the exponential map $\exp_x$ is a diffeomorphism from a subset $U_x \subseteq T_xM$ onto U.

By definition, $\exp_x(\xi) = \gamma_\xi(1)$, where γ_ξ is a g-geodesic in M such that $\left(\frac{d}{dt}\gamma_\xi\right)_{t=0} = \xi$. We may also "lift" $\exp_x$ to v by defining $\widehat{\exp}_v : U_x \to P_{\Gamma,u}$, where $\xi \mapsto (\hat{\gamma}_\xi)_v(1)$. Similarly we may define $\exp_{\Psi(x)} : U'_{\Psi(x)} \to P'_{\Gamma',u'}$ on M' using g', in which case $U'_{\Psi(x)} = \Psi_*[U_x]$, and for any $\xi' \in U'_{\Psi(x)}$,

$$\exp_{\Psi(x)}(\xi') = \gamma_{\xi'}(1) = \Psi \circ \gamma_{\Psi^*(\xi')}(1) = \Psi \circ \exp_x(\Psi^*(\xi'))$$

since $g' = \Psi_*(g)$. (Recall that since Ψ is a diffeomorphism, we may define the pullback of vectors as $\Psi^* = (\Psi^{-1})^*$.) We also get that

$$\begin{aligned}
\widehat{\exp}_{f(v)}(\xi') &= (\hat{\gamma}_{\xi'})_{f(v)}(1) = (\widehat{\Psi \circ \gamma_{\Psi^*(\xi')}})_{f(v)} \\
&= T'_{\Gamma',\Psi\circ(\gamma_{\Psi^*(\xi')}\bullet\beta_v)}(u') \\
&= f(T_{\Gamma,\gamma_{\Psi^*(\xi')}\bullet\beta_v}(u)) \\
&= f \circ \widehat{\exp}_v(\Psi^*(\xi')).
\end{aligned}$$

Now define a smooth local section $\sigma : U \to P_{\Gamma,u}$ as $\sigma = \widehat{\exp}_v \circ \exp_x^{-1}$. Then

$$\sigma' = f \circ \sigma \circ \Psi^{-1} = f \circ \widehat{\exp}_v \circ \exp_x^{-1} \circ \Psi^{-1} = \widehat{\exp}_{f(v)} \circ \Psi_* \circ \exp_x^{-1} \circ \Psi^{-1}$$

is a smooth local section of $P'_{\Gamma',u'}$. Now let $\eta : U \times \Phi_{\Gamma,u} \to \tilde{\pi}^{-1}[U]$ be a local trivialization of $P_{\Gamma,u}$ such that $\eta^{-1}[\sigma[U]] = U \times \{e_G\}$, and let $\eta' : \Psi[U] \times \Phi'_{\Gamma',u'} \to \tilde{\pi}'^{-1}[\Psi[U]]$ be a local trivialization of $P'_{\Gamma',u'}$ such that $\eta^{-1}[\sigma'[\Psi[U]]] = \Psi[U] \times \{e_{G'}\}$. Then we can write f locally as

$$f_{\restriction U} = \eta' \circ (\Psi \times \phi) \circ \eta^{-1}$$

since for all $w \in \tilde{\pi}^{-1}[U]$, we can write $w = yg$ for some $y \in \sigma[U]$. Then

$$\begin{aligned}
\eta' \circ (\Psi \circ \phi) \circ \eta^{-1}(w) &= \eta' \circ (\Psi \circ \phi)(\tilde{\pi}(w), g) \\
&= \eta'(\Psi \circ \tilde{\pi}(w), \phi(g)) \\
&= \eta'(\Psi \circ \tilde{\pi}(w), e_{G'})\phi(g) \\
&= \sigma'(\Psi \circ \tilde{\pi}(w))\phi(g) \\
&= f \circ \sigma \circ \Psi^{-1}(\Psi \circ \tilde{\pi}(w))\phi(g) \\
&= f \circ \sigma \circ \tilde{\pi}(w)\phi(g) \\
&= f(y)\phi(g) = f(w).
\end{aligned}$$

Since v was arbitrary, f is smooth everywhere. An analogous procedure can be performed for f^{-1}. □

We now prove the main result. Again, we restate it first for convenience.

Theorem 3.2. The Barrett construction functor $C_B : \mathbf{Hol} \to \mathbf{PC}$ forgets only structure.

Proof. Let $C_B : \mathbf{Hol} \to \mathbf{PC}$ be a functor that takes holonomy maps $H : L_x \to G$ on a manifold M to a principal bundle $G \to P \xrightarrow{\pi} M$ and principal connection Γ given by the Barrett reconstruction theorem—i.e., to a bundle and connection $(G \to P \xrightarrow{\pi} M, \Gamma)$ such that there exists a point $u \in \pi^{-1}[x]$ satisfying $H_{\Gamma,u} = H$—and takes a holonomy isomorphism $(\Phi, \underline{\alpha}, \phi)$ to the principal bundle isomorphism $(F, \Psi, \phi) : C_B(H_{\Gamma,u}) \to C_B(H'_{\Gamma',u'})$ given in the proof of Theorem 3.2. First, note that C_B clearly preserves holonomy data, and thus preserves empirical content in the required sense. We will first show that C_B is indeed a functor, and then show that C_B is one half of an equivalence, by showing it is full, faithful, and essentially surjective.

First, it is clear from the definition of F that $C_B((\Psi, \underline{\alpha}, \phi) : H \to H') = (F, \Psi, \phi) : C_B(H) \to C_B(H')$. It remains to show that $C(\mathrm{id}_H) = \mathrm{id}_{C_B(H)}$ and that $C_B(g \circ f) = C_B(g) \circ C_B(f)$ for any arrows $f : H \to H'$ and $g : H' \to H''$ of **Hol**. So let H be an arrow of **Hol**, suppose $C_B(H) = (G \to P \xrightarrow{\pi} M, \Gamma)$, and suppose $u \in \pi^{-1}[x]$ is such that $H_{\Gamma,u} = H$. Then $C_B(\mathrm{id}_H) = C({}_B(\mathrm{id}_M, \underline{\mathrm{id}_{\pi(u)}}, \mathrm{id}_G)) = (\mathrm{id}_P, \mathrm{id}_M, \mathrm{id}_G) = \mathrm{id}_{C_B(H)}$. Thus identities are preserved. Now let $(\Psi, \underline{\alpha}, \phi) : H \to H'$ and $(\Psi', \underline{\alpha}', \phi') : H' \to H''$ be isomorphisms of holonomy maps $H : L_x \to G$, $H' : L_{x'} \to G'$ and $H'' : L_{x''} \to G''$. Let (P, Γ), (P', Γ'), and (P'', Γ'') be the corresponding principal bundles and connections in the Barrett construction, and let $u \in \pi^{-1}[x]$, $u' \in \pi'^{-1}[x']$, and $u'' \in \pi''^{-1}[x'']$ be such that $H = H_{\Gamma,u}$, $H = H_{\Gamma',u'}$, and $H = H_{\Gamma'',u''}$, respectively. Then

$$
\begin{aligned}
&C_B((\Psi, \alpha, \phi) \circ (\Psi', \alpha', \phi')) : H \to H'' \\
&= C_B(\Psi' \circ \Psi, \alpha \bullet (\Psi^{-1} \circ \alpha'), \phi' \circ \phi) : H \to H'' \\
&= (F'', \Psi' \circ \Psi, \phi' \circ \phi) : C_B(H) \to C_B(H'')
\end{aligned}
$$

Where for $v \in P$, if $v = xg$ for $x \in P_{\Gamma,u}$, $g \in G$, then

$$\begin{aligned} F''(v) &= T_{\Gamma'',\Psi^{-1}\circ\Psi(\alpha\bullet(\Psi^{-1}\circ\alpha'))}(u'')(\phi'\circ\phi)(g) \\ &= F'(f(x)\phi(g)) = (f'\circ f)(x)(\phi'\circ\phi)(g) \\ &= F'\circ F(v) \end{aligned}$$

We now show that C_B is faithful, and essentially surjective, but not full.

To see that C_B is faithful, suppose there are two holonomy isomorphisms $(\Psi, \underline{\alpha}, \phi)$ and $(\Psi', \underline{\alpha}', \phi') : H \to H'$ which are such that $C_B(\Psi, \underline{\alpha}, \phi) = C_B(\Psi', \underline{\alpha}', \phi') = (F, \Psi'', \phi'')$. Then by the definition of C_B on arrows, $\Psi = \Psi' = \Psi''$ and $\phi = \phi' = \phi''$. Thus for all $\gamma \in L_x$,

$$H'(\Psi\circ(\alpha^{-1}\bullet\gamma\bullet\alpha)) = \phi\circ H(\gamma) = \phi'\circ H(\gamma) = H'(\Psi\circ(\alpha'^{-1}\bullet\gamma\bullet\alpha'))$$

Thus $\underline{\alpha} = \underline{\alpha}'$, and so $(\Psi, \underline{\alpha}, \phi) = (\Psi', \underline{\alpha}', \phi')$ and C_B is faithful.

To see that C_B is essentially surjective, let $G \to P \overset{\pi}{\to} M$ be a principal bundle with connection Γ, $(P, \Gamma) \in \mathbf{PC}$. Then $C_B(H_{\Gamma,u}) = (P, \Gamma)$ for some $u \in P$. So C_B is essentially surjective.

To see that C_B is *not* full, consider the holonomy map $H : L_x \to G$ on a contractible manifold M defined by $\gamma \mapsto e_G \in \mathbf{Hol}$ for some $x \in M$, where e_G is the identity in G. Then $C_B(H) = (P, \Gamma)$ for some principal G-bundle over M and flat connection Γ. Since this holonomy map is flat and M is contractible, there are no non-trivial equivalence classes of curves α (under the equivalence relation $\sim$ defined in the paper), so all elements of $\mathbf{Hom_{Hol}}(H, H)$ have the form $(\Psi, \underline{\mathrm{id}_x}, \phi)$ for some spacetime diffeomorphism Ψ and some Lie group automorphism $\phi : G \to G$. $C_B(\Psi, \underline{\mathrm{id}_x}, \phi) = (F, \Psi, \phi) \in \mathbf{Hom_{PC}}((P, \Gamma), (P, \Gamma))$, where F is constructed by the procedure in Theorem 1. Let g be any non-identity element of G. Then $(g \cdot F, \Psi, \phi)$ is a distinct element of $\mathbf{Hom_{PC}}((P, \Gamma), (P, \Gamma))$ from (F, Ψ, ϕ). However, $(g \cdot$

$F, \Psi, \phi) \neq C_B(\Psi, \underline{\alpha}, \phi)$ for any $(\Psi, \underline{\alpha}, \phi) \in \mathbf{Hom_{Hol}}(H, H)$, since elements of $\mathbf{Hom_{Hol}}(H, H)$ are uniquely defined by Ψ and ϕ. Thus C_B restricted to the automorphisms of H is not surjective. □

3.5 Wilson Loops

Healey (2007, p. 73) also discusses Wilson loops, or the traces of holonomy maps[1], as a possible alternate formalism for Yang-Mills theory that lacks the "excess structure" of the principal bundle formulation. Healey here appears to follow Gambini and Pullin (2000) and other mathematical physicists in claiming that the method introduced by Giles (1981) to reconstruct holonomy maps from Wilson loops modulo similarity transformations proves that Wilson loops contain all of the gauge invariant physical content of a Yang-Mills theory. The idea is that since the trace of a Lie group matrix will be invariant under conjugation by other group elements (*i.e.*, transformations of the form $a \mapsto gag^{-1}$ for $a, g \in G$), and since holonomy maps defined at different base points are related by conjugation, moving from holonomies to Wilson loops eliminates the excess baggage that comes with the choice of base-point, and thereby expresses only the parts of the theory that are independent of such a choice of representation.

Unfortunately, it is not clear that the proofs provided in Giles (1981) prove that holonomies can be reconstructed from Wilson loops in full generality. Giles provides a method for constructing matrices with traces corresponding to Wilson loop values to associate with equivalence classes of curves. However, Giles notes that he has not imposed sufficient conditions for admissible generalized Wilson loops so as to guarantee that the reconstruction has all of the desirable properties of a holonomies. It is nonetheless a promising direction for research.

[1]The trace is the sum of diagonal elements of matrices corresponding to the holonomy values (elements of the group G) represented on some vector space.

This lack specificity for the conditions on Wilson loops for the Giles reconstruction prevents us from adequately defining a category **Wil** of Wilson loops and comparing it to **PC** and **Hol**. We can nonetheless still sketch what such a TCM analysis might look like. Analogously to **Hol**, the objects of **Wil** would be a generalized Wilson loop maps $W : L \to \mathbf{C}$ (where loops no longer require a fixed base point x) satisfying certain properties which would allow them to be thought of as the traces of some holonomy map. Whatever the isomorphisms are, they should be such that isomorphic Wilson loop maps can be thought of as the traces of the same holonomy maps under possibly different matrix representations.

The functor $F_\rho : \mathbf{Hol} \to \mathbf{Wil}$ takes the trace of holonomy values according to some vector space representation $\rho : G \to \mathrm{GL}(V)$. For the Giles reconstruction $C_G : \mathbf{Wil} \to \mathbf{Hol}$ to be a proper reconstruction, it would need to be faithful and essentially surjective, and conversely F_ρ would need to be full and essentially surjective. But F_ρ might forget *stuff*. Indeed, we might expect it to. The motivation for moving to Wilson loops, after all, was to eliminate distinctions between conjugate holonomy maps. If F_ρ does so, it will identify maps related by conjugation, including those related by a base-point transformation. So at the very least, we would expect base-point transformations given by distinct curves—and thus distinct holonomy isomorphisms—to map to the same Wilson loop isomorphism under F_ρ. As with $C_B : \mathbf{Hol} \to \mathbf{PC}$, the "structure" that C_G forgets is indicated by what is required in order to define an inverse. A functor for $F_\rho : \mathbf{Hol} \to \mathbf{Wil}$ will take holonomies to their traces in some matrix representation ρ of G, and so requires fixing a preferred representation. This can be expressed more abstractly as defining a *character*[2] on G, but it still involves specifying an additional structure of some sort.

While the functor $C_G : \mathbf{Wil} \to \mathbf{Hol}$ might provide us a sense in which any Wilson loop map $W : L_x \to \mathbf{C}$ allows us to construct a holonomy map $H : L_x \to G$, it does not allow us to reconstruct all of the gauge-invariant content of the holonomy $H(\gamma)$ of a *specific curve* γ from

[2]See Brocker and tom Dieck (1985) chapter 2 for details.

its Wilson loop $W(\gamma)$. If the holonomies of two loops are related by non-trivial conjugation by a Lie group element, then their holonomies will be related by non-trivial conjugation by *some* Lie group element in *any* representation of the holonomies at any base-point.

In particular, suppose that for a holonomy model $H : L_x \to G$, I have two loops γ_1 and $\gamma_2 \in L_x$ which are such that $H(\gamma_1) = g_1$ and $H(\gamma_2) = g_2 = ag_1a^{-1} = aH(\gamma_1)a^{-1}$ for some $g_1, g_2, a \in G$, and $g_2 \neq g_1$. Let H' be a distinct but isomorphic holonomy model on the same base space which is such that $H'(\gamma) = bH(\gamma)b^{-1}$ for some $b \in G$ for all $\gamma \in L_x$. In other words, H and H' can be thought of as corresponding to the same principal bundle model with a different choice of base point. Then $H'(\gamma_2) = bH(\gamma_2)b^{-1} = baH(\gamma_1)a^{-1}b^{-1} \neq H'(\gamma_2)$. Thus if holonomies of γ_1 and γ_2 are related by some non-trivial conjugation in some holonomy model, they are so in any holonomy model related to it by change of base point.

Another worry is that the trace of a Lie group element is not in general invariant under different choices of representation ρ of the Lie group G. This undermines a crucial role of principal bundles with connections in coordinating the influence of the Yang-Mills field on other fields which take values in different vector spaces (see Weatherall (2015) for details). Performing this function requires the ability to interpret representations of G on different vector spaces corresponding to different space-time fields. Elements of these representations will have different traces, and it is thus unclear how a Wilson loop picture could play this unifying role of the principal bundle with connection.

In sum, an adequate presentation of Yang-Mills theory in terms of Wilson loops is not fully developed, but it is promising and worthy of pursuit. It also would take the apparent non-locality of holonomies a step further, attributing properties not just to closed curves but to the behavior of curves when conjugated with other curves. This indicates the presence of some interesting, possibly generalizable trade-offs between an increasingly algebraic picture with fewer distinct but equivalent models, and the ability to describe systems locally in space.

$$
\begin{array}{ccc}
\mathbf{PC} & \xleftrightarrow{\cong} & \mathbf{Trans} \\
{\scriptstyle C_B}\big\uparrow & & \\
\mathbf{Hol} & & \\
{\scriptstyle C_G}\big\uparrow & & \\
\mathbf{Wil} & &
\end{array}
$$

Figure 3.4: Yang-Mills groupoids and the functors between them.

3.6 Discussion

The goal of this chapter was to apply TCM to elucidate the relative structural content of various Yang-Mills formalisms. Doing so involved proving novel results about the relationships between principal bundles and holonomy maps, sharpening the picture Barrett (1991) provides with his reconstruction theorem. Schreiber and Waldorf (2009) have similarly proved an equivalence between **PC** and a category **Trans** of generalized parallel transport functors. Our work expands the menagerie of Yang-Mills categories to include **Hol** and **Wil**, generating a TCM portrait illustrated in figure 3.4.

The Barrett construction functor $C_B : \mathbf{Hol} \to \mathbf{PC}$ and the (not completely defined) Giles functor $C_G : \mathbf{Wil} \to \mathbf{Hol}$ both simultaneously *forget structure* and *add stuff*. The structure that each forgets is indicated by what is required to construct an inverse. For $C_B : \mathbf{Hol} \to \mathbf{PC}$, constructing an inverse involves choosing a preferred point u in each principal bundle object. For $C_G : \mathbf{Wil} \to \mathbf{Hol}$, constructing an inverse would require (at least) choosing a preferred matrix representation of the Lie group. C_B can be thought of as adding the non-holonomy sub-bundle "stuff" of a principal bundle, while C_G adds base points to curves on spacetime.

This PSS story does not straightforwardly refute Healey's claim that holonomies are more parsimonious than principal bundles. Rather, it prompts more specificity regarding what parsimony considerations are supposed to refer to, and in what ways they appear in this case. It is not the case, for example, that principal bundles have "more structure" than holonomies in the form of gauge, since the functor $D_u : \mathbf{PC} \to \mathbf{Hol}$ demonstrates how gauge transformations are similarly present in the holonomy models. The Barrett construction functor $C_B : \mathbf{Hol} \to \mathbf{PC}$ additionally reveals how a principal bundle (at least, the holonomy sub-bundle), is *structurally* present in the holonomy formalism, even though it is not explicitly stated. Moreover, holonomies are in a sense more "structurally presumptuous": moving from **PC** to **Hol** requires picking out a preferred element u of each principal bundle in order to define $D_u : \mathbf{PC} \to \mathbf{Hol}$.

Healey might nonetheless claim that the sense of parsimony he was invoking is more adequately captured by the fact that $D_u : \mathbf{PC} \to \mathbf{Hol}$ *forgets stuff*. Indeed, as Bradley and Weatherall (2019) argue, philosophers of physics do sometimes implicitly invoke "stuff-forgetting" as a way of increasing parsimony, as in Nguyen et al. (2017). As argued in chapter 1, however, PSS analysis is less clear and more perspectival in cases involving split functors like C_B. From this vantage, however, the even more structured but less stuffy Wilson loops may be more appealing than holonomies. One interesting upshot of the analysis of Wilson loops in section 3.5 is that it reveals how quite a lot can be said about a PSS relationship without precisely defining the categories or functors involved. Category theoretic tools can be employed to express the "shape" of concepts that have not been quite pinned down, and reveal the functional role that they play.

Figure 3.4 tells a story about various formalisms that can be used to represent states of affairs according to (classical) Yang-Mills theory. This general picture relating different formalisms and expressing their trade-offs is the most significant upshot of this project. Physicists do not typically think of classical Yang-Mills theories as themselves physically meaningful prior to being quantized. But the right way to quantize Yang-Mills theories is still an open problem. In fact, the Clay Mathematics institute offers a $1,000,000 prize for a proof of the existence of a model of quantum Yang-Mills theory satisfying minimal conditions for empirical adequacy. Quantization procedures start with some classical presentation of Yang-Mills theory, and and can be expressed functorially.[3] The functorial relationships between classical formalisms described here might prove useful in this cutting edge endeavor.

[3] See, for example, Rejzner (2016).

Chapter 4

Topological Data Analysis

4.1 Introduction

Data scientists take large quantities of noisy measurements and transform them into tractable, qualitative descriptions of the phenomena being measured. While it frequently involves statistical methods, the burgeoning field of data science distinguishes itself from statistics by branching out to a wider range of methods from mathematics and computer science. One such distinctly non-statistical method of growing popularity is topological data analysis (TDA). Topology is the study of the properties of shapes that are invariant under continuous deformations, such as stretching, twisting, bending, or re-scaling, but not tearing or gluing. TDA aims to identify the essential "structure" of a data set as it "appears" in an abstract space of measurement outcomes. This paper is an attempt to understand how it achieves this, drawing on the TCM perspective developed in chapter 1.

Data scientists themselves invoke category theory to justify TDA's core methods. At the heart of TDA is the concept of *homology*, an abstract mathematical interpretation of "hole" structure. Homology exhibits the category theoretic property of *functoriality*, meaning it is defined not only on models but on structure-preserving functions between them. There are ubiquitous hints in the TDA literature that its practitioners consider the functoriality of homology to be central to its utility in application, but this maxim does not appear to be explored in much depth. This discussion of TDA aims to provide insight into how scientists are and can be motivated by category TCM-esque considerations, and the ways in which these motivations do and do not align with those of philosophers of physics employing the same tools. I argue that the utility of category theoretic methods to researchers in this context is rooted in the particular geometric nature of the mathematical models. The category theoretic framework helps to connect topological models, which have straightforward physical interpretations, with algebraic models, which are more abstract but easier to process computationally.

In section 4.2 I describe TDA in detail. Section 4.3 discusses the role of category theory in TDA, and section 4.4 examines the role of spatial reasoning in TDA, and how it can give us insight into the connections between the data scientists' and philosophers' notions of "structure." I conclude by reflecting on how this discussion can enrich the account of TCM presented in chapter 1.

One curious feature of the TCM literature in philosophy of physics is its focus on *groupoids*, or categories with only *isomorphisms*—invertible, fully structure-preserving maps between models. This is in stark contrast to the categories that appear in TDA, where maps between models that "forget" some structure play a central role. This indicates that groupoids may be insufficient for characterizing the representational capacities of formal models, and only reveal a fraction of the potential insight TCM has to offer. Philosophers of physics have largely relied on the notion that the groupoid tells us when two models represent the same system. In TDA, we see how additional arrows can give insight into the *internal* character of models, expressing not just the fact of sameness of models, but which *parts* of a model in one formalism correspond to which in another, as well as cross-model identification of parts.

4.2 Topological data analysis

The phrase "topological data analysis" is used to refer to a variety of data science practices that use tools from algebraic topology to make inferences about the "shape" of data clouds as they appear in the "space" of possible observations. For now, the term *data* refers to a set of real vectors corresponding to a series of observations. This is an adequate definition for capturing natural language use of the term, but one might object that it does not necessarily capture what data *is*. One of the goals of TDA is to circumvent some of the arbitrariness involved in presenting data as real vectors. A *data cloud* can thus be thought of as a visual representation of this set of vectors as "points" in a (high dimensional generalization of) space. But in what space? The abstract "space" where data lives is generally some form of *metric space*, or set X of points (including at least the data points) together with a notion of "distance" $d(\, , \,)$ between the points. For example, I may have data about the weights of each of a large number of potatoes. The distance between these data points would just be the pairwise difference in weight between two potatoes according to a fixed unit, such as pounds.

An advantage of looking at geometric properties is that it moves away from the full vector space, which includes a choice of which value counts as "0" as well as choice of coordinate system. For the philosopher of physics, this might be sufficient—indeed, complaints about arbitrary choices of coordinates are common in this literature. But further complications arise outside of physics. For example, in social science, subjects may be asked qualitative questions ("How do you feel on a scale of 1-10?") in which the full structure of the metric space is not so meaningful as the order, or whether the number is above or below a threshold. Carlsson (2009) further notes that, especially in computational biology, "notions of distance are constructed using some intuitively attractive measures of similarity (such as BLAST scores or their relatives), but it is far from clear how much significance to attach to the actual distances" (p. 256).

A characteristic problem of analyzing large data sets is deciding how to combine many different types of measurements into a shared metric space. I can also add information about the length, color, number of eyes, etc. for each potato, creating an n-dimensional space, where n is the number of potato attributes. The "distance" between two data points is now some combination of the distances given by weights, lengths, color, etc. But how should the notions of distance given by each variable combine into "distance" in the total space of possible variable values? The "standard" way of aggregating one-dimensional metrics into a shared metric space is to imagine each metric as an axis in an n-dimensional Cartesian grid, with distance given by the Cartesian distance as follows. Let $x = (x_1, ..., x_n)$ and $y = (y_1, ..., y_n)$ be two sets of potato measurements. Then $d(x, y) = \sqrt{(x_1 - y_1)^2 + ... + (x_n - y_n)^2}$. Setting aside the fact that there are other viable options for constructing distances from these values, notice that this expression does not include units. Should weight be presented in pounds or tons? Of course we know how to translate between these two units, and we consider the choice more of notational convenience than theoretically meaningful. But if we are looking to the "shape" of data for information about the system being measured, the data cloud will look much more "flat" if we use tons rather than pounds. It is thus desirable to consider properties of the data cloud that do not depend on the particular choice of metric space or unit, but which are shared by a variety of plausible modeling choices.

Such considerations motivate the use of *topological*, as opposed to geometric methods. Topology is the mathematical field that studies properties of shapes that remain constant under stretching, twisting, or otherwise deforming. (Topology nicknamed "rubber-sheet geometry" for this reason.) Topologists attend to more general features of metric spaces that would be present under different modeling assumptions, called *topological invariants*. Since data sets are finite, although they may suggest some underlying shape, they likely will not do so uniquely. This is the standard curve-fitting problem in higher dimensions: for any discrete set of points, there are an infinite number of continuous curves (or shapes) that contain (or approximate) the locations of those points. As with the curve-fitting problem, external considerations guide the choice of continuous object, rather than just the bare, uninterpreted set of data points. One may have a priori reasons to expect that the "right" curve is quadratic, for example, or that the modeling goal should be to minimize mean-squared error.

4.2.1 Clusters

The simplest example of TDA, and the one most broadly used by data scientists generally, is *cluster analysis*. The idea behind cluster analysis is to ask: do my data points naturally divide into sub-categories of data points more similar to one another than the overall space? Such a situation indicates that there is some non-trivial structure underlying the data associated with such groupings, which one may interpret as "natural kinds" in the space. Cluster analysis is in this way closely related to regression analysis—clusters point towards a correlation among variables, one of the main "signals" data scientists hope to read off of large data sets.

Sometimes, external considerations about the type of data under consideration can influence how one choses to carve a data set into clusters. Even in the absence of such guidance, natural clusters may be easily "seen" when the data is graphed. With larger and higher dimensional data sets to analyze, these heuristics are less useful, and data scientists would prefer a principled algorithmic approach to clustering. This would amount to a function that takes metric spaces (X, d)—here understood as data sets $X = \{x_1, ..., x_n\}$ with a notion of "distance" $d(x_i, x_j)$—as inputs, and outputs *partitions* of that data into clusters of data points that are "close together." There are two major barriers to creating such an algorithm.

The first is an impossibility theorem from (Kleinberg, 2003) showing that there is no non-trivial clustering algorithm that simultaneously satisfies the following seemingly reasonable minimal requirements:

1. *Scale invariance:* if two metrics on the same data set differ by a constant multiple, they output the same clusters;

2. *Surjectivity:* for all possible partitions of X, there is some metric on X for which the algorithm outputs this partition;

3. *Consistency:* If distances are reduced between points in the same cluster, and increased between points in distinct clusters, the output is the same partition.

This impossibility theorem is sobering for methodological purists hoping to precisely and naturally read off clusters from metric spaces of data. The pragmatic data scientist may not be perturbed, however, since the impossibility of a complete clustering algorithm does not preclude useful heuristic clustering methods for cases of interest. But even the pragmatist may be dismayed by the fact that for large, high-dimensional spaces, it is impractical to check heuristics against our intuitions about "good" clustering algorithms in order to be assured of the consistency of the analysis.

To make matters worse, attempts to read shapes from data clouds may give different results when looking at the data at different "resolutions." Again, external knowledge about the system may indicate which resolution is of interest, but again this defeats the goal of developing tools to analyze large data sets that we do not comprehend. In section 4.3.3, we will see how re-framing clustering as a functor rather than a function resolves these issues.

4.2.2 Constructing Shapes

The most common method to construct a shape from a data cloud is roughly as follows. Enclose each data point in a "ball" of radius ε centered on that point. As ε gets larger, the cloud will cease to look like isolated points and start to gain shape. Once it gets too large, though, we are left with a single shapeless blob. We use this idea to construct a *simplicial* *complex*, beginning with the data points as vertices.[1] Where 2 balls intersect, we add an *edge* between them. When 3 balls intersect, we add a *face* enclosed by the three edges. When 4 intersect, we create a *cell*, a triangular prism enclosed by the four edges. This process continues, creating higher dimensional *n-faces* where $n + 1$ balls intersect. The result is called a *Čech complex*.[2]

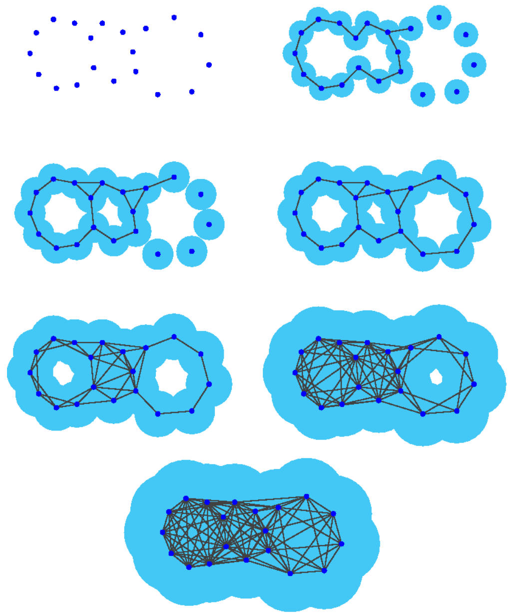


Figure 4.1: Constructing a Čech complex as ε increases, from Bubenik (2015).

Setting aside for the moment the problem of selecting the right resolution (which here is captured by the choice of ε), this is an intuitively plausible way to construct a discrete shape from a data cloud. A clustering can be read off of a Čech complex by grouping data points according to whether they are connected in a single component of the complex. This may be complicated by the presence of noise—a single anomalous data point might connect otherwise robustly distinct clusters. This can be dealt with by either looking at only regions that are highly connected, or avoided altogether by filtering and "cleaning" the data prior to analysis.

[1]See Hatcher (2002) section 2.1 for a precise definition of a simplicial complex.

[2]In practice, TDA employs a more computationally tractable approximation thereof, called a *witness complex*. See Carlsson (2009) section 2 for details.

While this seems like a reasonable way to cluster data, it nonetheless follows from the impossibility theorem from the previous subsection that this method lacks the desirable properties listed there. As we continue, we will see that this is not as much of a problem as it initially appears. The reader should begin to cogitate on how we will ultimately use category theory to get around the challenge posed by the theorem.

4.2.3 Holes and voids

To those familiar with algebraic topology, identifying the clusters of a simplicial complex appears to be a special case of a more general phenomenon of *homology*. Homology refers to a method of classifying shapes by looking at how many "holes" the shape has. No matter how much you stretch and twist it, a circle will always have a "hole" in it, a sphere will always have a *void* or *cavity*, an innertube will always have the "donut hole" as well as a void in the interior that inflates.

In looking at the connected components of a Čech complex, we are considering the H_0-*homology* of the complex (considered as a topological space). We can similarly attend to the H_1-*homology* of the complex by looking for "holes," or the H_2-*homology* by looking at "cells," and so on to higher dimensions with less intuitive interpretations.[3]

Example 4.1 (Cosmology)**.** van de Weygaert et al. (2011) study the homology of density level sets of an ensemble of randomly generated cosmic mass distributions. They analyze the evolution of H_1, H_2, and H_3-homology over time in n-body simulations, revealing characteristic patterns of different dark energy models. They show how homology can track cosmological structures of independent interest to physicists, such as matter power spectra and non-Gaussianity in the primordial density field.

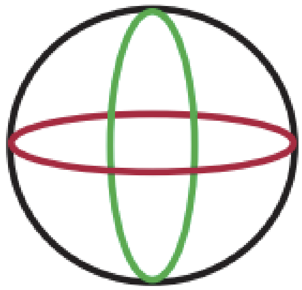
[3]I am speaking very abstractly and non-rigorously here. I will go into a bit more detail later on, but for a complete account of homology refer to Hatcher (2002), or Ghrist (2014) for an account catered specifically to TDA.

See below for a more detailed example, or feel free to skip ahead to the next section.

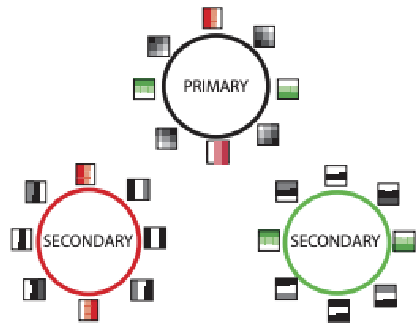
Example 4.2 (Natural Image classification)**.** Carlsson et al. (2008) apply TDA to make precise the qualitative features observed in Lee et al. (2003). The latter authors sampled a large corpus of 3×3 pixel patches of gray-scale photographs, presented as 9-tuples of real numbers corresponding to the gray-scale value at each pixel in the patch, yielding a data cloud in $\mathbb{R}^9$. They normalized this data set to identify patches that relate to one another by a shift in "brightness" or "contrast"—i.e., a function performed in image post-processing. This projected the data onto a 7 dimensional sphere in $\mathbb{R}^8$. The authors noted that the points were scattered across the sphere, but with highly varying density. They noted in particular that the data was largely concentrated around an annulus.

Carlsson et al. (2008) show how this can be summarized by saying that natural images have certain characteristic homology. The data is first "cleaned" by restricting attention to data points that exist in higher density regions—this is done for varying definitions of density via varying the k in a k-means filter. They show that for large k, i.e. a high threshold for density, for a robust range of ε choices, the resulting Čech complex consisted in a single, persistent annulus. Data points in this annulus corresponded to patches that transition from dark to light, with position along the annulus corresponding to "orientation" of the transition.

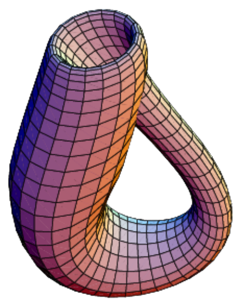
Relaxing the threshold for density, the resulting characteristic shape included the same annulus, along with two secondary, perpendicular annuli. The secondary annuli roughly corresponded to patches consisting of columns (respectively horizontal) of similar shades that do not transition smoothly—i.e. light-dark-medium, light-dark-light, etc. Carlsson et. al. noted that this 3-circle model embeds naturally in a Klein bottle. They construct an algorithm to modify data such that for natural image data, the resulting H_0, H_1, and H_2-homologies correspond to those of a Klein bottle—1:2:1, while randomly generated data produces no notable homological features.



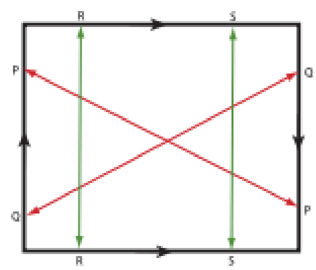
(a) 3 circle model



(b) The data



(c) Klein bottle



(d) Embedding

Figure 4.2: From Carlsson (2009).

4.2.4 Persistence

The motivating idea behind the construction of a Čech complex is that we can imagine data as being uniformly sampled (with noise) from some underlying “shape” in the metric state space, and we can use these data points to infer the global structure of the “object” we are sampling from. The more samples we look at, the more accurate our picture of the shape will be. For sufficiently small ε-balls, the complex will not have any more structure than the bare data set. Similarly, when the balls get too large, there is nothing more to look at than a giant blob. The “right” choice of ε is at some intermediate size, but how should it be chosen? If we chose an ε that is too small, we will get a shape with a lot more holes, disconnected components, etc., than we think are meaningful. In other words, we retain some of the noisy features of the data cloud that we were trying to eliminate. But we risk going to far, and making ε large enough to obscure both noise *and* meaningful information from the data.

A natural way to solve this problem is to look at many different choices of ε, and use external considerations to decide which gives the best resolution of the data shape. Two more problems arise when we do this, though. For one, the whole point of data analysis is to simplify and compress information about a system, and having a variety of different models we can choose from does not simplify matters. Second, there may be different features that arise at different resolutions that are equally significant, and this multi-level picture can get lost if we have to choose a single model among the many possibilities. For example, data may be dense in some regions but sparse in others, where relevant shapes require larger ε-balls to be "seen".

The key insight that unlocked the power of TDA was the idea of "topological persistence," introduced to data analysis in (Edelsbrunner et al., 2002). Briefly: instead of picking a particular resolution to look at, we look at them all, but take advantage of a trick from algebraic topology to connect complexes at different scales in a sophisticated and efficient way (read: *functorially*). The result is the association of a data cloud with a *persistence module* that encodes how the cloud changes structurally as ε increases. Homology is then computed for these modules, and the result is typically expressed as a *homological barcode*, as in figure 4.3. The "bars" begin when a feature is "born" and end when it "dies." Short intervals in barcodes are often attributed to either measurement noise or inadequate sampling, whereas long, "persistent" bars are thought to reveal real geometric features of the space being sampled from.

This construction is enabled by a structure theorem of Crawley-Boevey (2015), demonstrating that persistent modules can be uniquely represented as a direct sum of *interval modules*. Not only is this decomposition more computationally tractable to analyze than (sets of) complexes, but the barcode itself provides a visual summary of behavior as ε increases. When the number of features is large, data analysts will also sometime use *persistence diagrams* instead of barcodes. These diagrams plot features on a birth-death axis. See figure 4.4 for

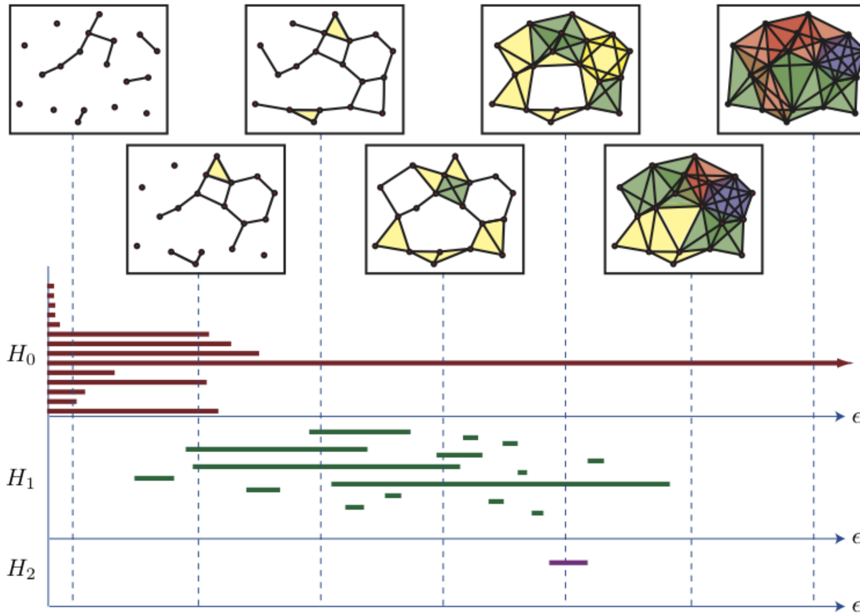


Figure 4.3: Example of a homological barcode, from Ghrist (2008).

a diagram of voids—H_2-homological features—in a cosmological model from example 4.1. Dots on the diagonal indicate voids that die quickly after birth, and those farther away are more persistent.

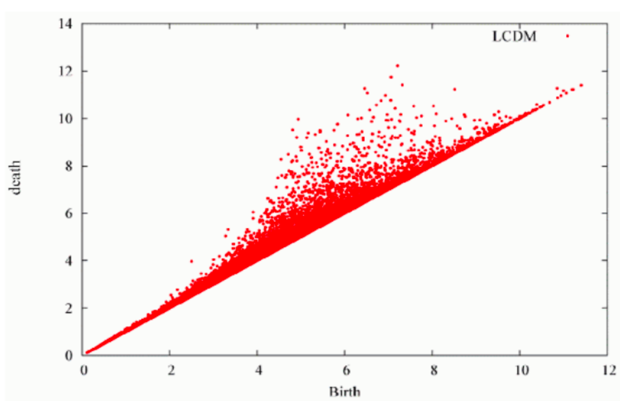


Figure 4.4: Birth-death diagram of voids in a cosmological model (van de Weygaert et al., 2011).

4.2.5 Stability

One way to interpret ε is as a modeling parameter, corresponding to the resolution or scale we use to construct a shape from the data cloud. The persistent features of a Čech complex are those that are *stable*, or robust under perturbations of the parameter value. Longer bars in barcodes represent features that appear for a wider range of ε values, indicating that these features are robust and unlikely to constitute mere noise. Cohen-Steiner et al. (2007) made this precise by proving that for a large class of constructions (including Čech complexes), persistence diagrams are *stable*, meaning that small perturbations of the initial data set result in correspondingly small changes in the resulting persistence diagram.

We can use this same method to consider stability across other indexing parameters as well at fixed resolution, as in the following example.

Example 4.3 (Arteries)**.** Bendich et al. (2016) employ topological data analysis to study the structure of arteries in the human brain. They uniformly sample a large number of points from a blood vessel diagram (weighted by thickness of vessel), and construct a Čech complex from this data cloud, analyzing the H_0 and H_1 persistence diagrams over the growing size of ε-balls in the Čech complex. They look at persistent H_0 over a stack of "horizontal slices" of the artery diagram.

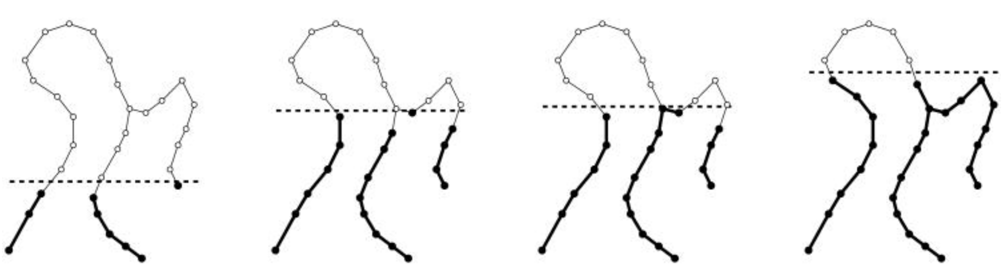


Figure 4.5: Horizontal slices of the artery diagram, from Bendich et al. (2016).

The authors found significant correlation between certain features of these homological barcodes and the age and sex of the subjects, with the age correlation a significant improvement over previous attempts at analyzing similar data. For example, older brains tended to have the longest bars in the latter barcodes.

In this example, persistence is indexed over the parameter of height. One can also analyze persistence of homological features over time.

Example 4.4 (Time-series data)**.** (Perea and Harer, 2015) demonstrate that persistent H_1-homology over time can be used to detect periodicity in time-series data by embedding it into a higher dimensional space. Note that in the absence of such an embedding, time series data displays no "loops" (since prior points in time are never revisited), so as it stands, it is not conducive to analysis of homology. It is fairly common for data analysts to modify their data to match their methods in this way, rather than the other way around.

We can thus understand persistence modules as assembling a sequence of $(n-1)$-dimensional models in a sequence indexed by an n^{th} parameter, such as resolution or time. Dimensionality reduction is a common feature of data analysis techniques. Data often comes in the form of large vectors, and the goal is often to *compress* them—express as much of the original information as possible with in as few dimensions as possible. This amounts to selecting features or parameters of interest and suppressing the rest in order to highlight general patterns. Reducing data models to 2-3 dimensions also makes them more visualizable, making them more useful to researchers to observe patterns, as well as easier to communicate to the public. Persistence modules provide the benefits of low dimensional visualizability without throwing away the information in the extra dimensions.

Summary

The general procedure for determining persistent homology is as follows.

1. Generate a sequence of shapes (CW-complexes) from the data cloud.

2. Transform the sequence into a *persistence module* indexed by a parameter such as resolution or time.

3. Construct a visual summary of the persistence module as a barcode or diagram.

1 and 3 are straightforwardly motivated—1 from the intuitive geometric interpretation of data as (noisily) sampled from some underlying shape, and 3 from the Crawley-Boevey structure theorem. In the next section, we how the move in 2 is motivated by the existence of a *functor* between a category of topological spaces of complexes constructed from data clouds on the one hand, and the category of homological algebras on the other. In section 4.4 I will argue that this functor is what transfers the epistemic value we grant to the shapes in 1 to the barcodes we construct in 3.

4.3 Functoriality in TDA

Whichever method we use to give shape to our data cloud, the result is a topological space. More specifically, it is a (finitely generated) *CW-complex*: a particularly "well-behaved" topological space that is constructed by "gluing" n-disks along their boundary $(n-1)$-spheres. Čech complexes are CW-complexes, as are all of the other constructions of figures from data clouds that we will consider here.

Homology is a general way of associating, to each of these shapes X built from a data cloud, a (finitely generated) Abelian *homology group* $H_n(X)$. For each group, $H_n(X)$ essentially characterizes how many "holes" are present in each dimension. $H_0(X)$ tracks the connected components, $H_1(X)$ tracks holes, $H_2(X)$ tracks cells, or the number of valves that would be required "inflate" the hollows of the shape. This extends to higher dimensions, but most TDA applications only look at these three, as these are the most spatially intuitive.

In order for persistence analysis to work, we need to be able to track shapes as they appear and disappear when ε increases. This is where the *functoriality* of homology comes in. Homology is not merely an assignment of a group to each complex that provides information about its shape. Homology is *functorial* in the sense that it comes equipped with a notion of how to translate maps between complexes into maps between groups while preserving all relevant topological information. This functoriality is inherited from the homology functor from the category of CW-complexes **CW** to the category of abelian groups **Ab** that lies at the heart of algebraic topology. The functoriality of homology enables us to do three important things, which are essential to its utility in analyzing data: identify local structures, connect complexes as parameters vary, and compare complexes constructed from different samples.

4.3.1 Locality

The homology group $H_n(C)$ of a complex C tells us how many "holes" it has, but it does not tell us *where* the holes are, or how big they are. This is to be expected—recall that while these complexes "live" in metric spaces, TDA looks at more general, topological rather than geometric features of them, which are preserved when the space is stretched or rotated. Nonetheless, topological spaces still have a (albeit weaker) notion of "nearness" associated with them. We can cover our topological space with "neighborhoods," and ask, relative to a particular cover, whether a "hole" is contained in a single neighborhood.

So, if there is a feature of interest, we can locate it in a neighborhood $U \subseteq C$ and think of this neighborhood as its own complex. We can then look at the inclusion map $\iota : U \to X$ that just acts as the identity on that neighborhood. Since homology is *functorial*, this induces a corresponding map $\iota_* : H_n(U) \to H_n(C)$, allowing us to track the n-dimensional "hole" in the group $H_n(C)$ as the image $\iota_*(H_n(U)) \subseteq H_n(C)$. (See Zomorodian, Afra and Carlsson, Gunnar (2008) for details on this localization method).

We can thus refer to a particular hole as it appears in the homology group, rather than referring to it spatially. But even more importantly, given a map $f : C \to D$ that identifies two complexes via their underlying metric space, we can ask whether the hole contained in U *persists* under the transformation f by seeing whether $f_*(\iota_*(H_n(U)))$ vanishes. This is what enables the use of homological barcodes to encode information about when holes form and disappear as a complex is constructed in stages by increasing ε. Each bar corresponds to a different hole, understood locally in this way.

4.3.2 Bootstrapping

The field of data science relies on the idea that sample data can in some situations be thought of as representative of the full statistical population from which the sample was taken. Such inferences from part to whole are of course not always warranted, so it is important to provide justification for such inferences when they are made. This is often done via external considerations ("I shuffled the deck really well!"), but there are also purely statistical methods of justification, called *bootstrapping* methods (Efron, 1979). The idea underlying bootstrapping is that inferences made from sample to population can be "modeled" as inferences from sub-samples to the full data set. The sampling error for such models is taken to indicate how much deviation the full data set may bear to the underlying population from which it was sampled.

There is a direct TDA analog to statistical bootstrapping. Given two samples from a data set, we can construct and compare sequences of complexes from each. It might not be sufficient, however, to merely observe whether pairs of complexes from each sample are qualitatively similar to one another. There is a possibility of a "false positive" indication that the two are qualitatively similar given that they are isomorphic to one another (have the same shape), but the isomorphism that relates them isn't the "right one." For example, at some value (or range) of ε, the Čech complexes might each display one hole, but they are "different" holes. Flat-footed comparisons are also prone to false negatives. If one sample is sparser than another, the "right" way to compare them might not be at the same resolution (choice of ε).

TDA resolves these issues by directing and constraining cross-sample comparisons to be consistent with the sense in which they are understood to be samples of the same underlying source. Both samples can be thought of as embedded in any subset $S \subseteq X$ that contains their union.[4] S thus provides a shared framework for comparison. Given elements $x_i \in H_n(C_{\varepsilon_i}(S_i))$ (i.e. n-dimensional holes), we can look at the inclusions $\iota_i : x_i \hookrightarrow H_n(C_\varepsilon(S))$ and check whether they map to the same features in the shared space.

4.3.3 Re-possibility theorem for clustering algorithms

The theorem in 4.2.1 shows that if one understands clustering algorithms as functions from (metric) data sets to partitions, then there is no possible algorithm that is consistent, scale invariant, and enables all possible partitions to be the result of some metric structure on a data set. To get around this problem, Carlsson and Mémoli (2008) re-define clustering algorithms as *functors* from a category **FinMet** of finite metric spaces to the category **Clust**. There are a few different reasonable choices for morphisms in this category, depending on the application. The objects of **Clust** are pairs $(\mathbb{X}, P^{\mathbb{X}})$ consisting of a finite metric space and a partition (clustering) thereof. The morphisms of **Clust** are functions $f : \mathbb{X} \to \mathbb{X}'$ that are such that $f^{-1}(P^{\mathbb{X}'})$ is a *refinement* of $P^{\mathbb{X}}$—i.e., cluster morphisms can merge clusters, but cannot break them up. Isomorphisms in **Clust** are then morphisms such that $f^{-1}(P^{\mathbb{X}'}) = P^{\mathbb{X}}$.

[4]Just the union may be sufficient, unless the samples come from different regions of the data cloud. This might be resolved by looking at the full data set, but again there are situations in which that is not feasible, leaving many situations to require some intermediary.

Conceiving of clustering as a functor requires that the morphisms in **FinMet** be "carried over" to **Clust**. Category theoretically, the requirement of "surjectivity" is replaced by "fullness"—every object in **Clust** must be *isomorphic* to a cluster that can be achieved by the algorithm. Which other properties you want your clustering algorithm to have will also inform your choice of morphisms for your metric space category. For example, the motivation behind the requirement of "consistency" is that a clustering functor should commute with distance non-increasing isometries. Thus the category of finite metric spaces should at least include such maps.

Where the category theoretic account of clustering really shines is in its explication of the "scale-invariance" criterion, which is the hardest to hash out in the previous paradigm. Now, instead of requiring that a clustering algorithm F be such that $F(\mathbb{X}) = F(\mathbb{X}')$ whenever $\mathbb{X}'$ is a rescaled version of $\mathbb{X}$, we instead require that for a rescaling $f : \mathbb{X} \to \mathbb{X}'$, there is a corresponding isomorphism $F(f) : F(\mathbb{X}) \to F(\mathbb{X}')$. Think about the clustering induced by Čech complexes. If you rescale your metric space, then the sense in which you should expect "the same" clustering to result (the motivation behind the scale invariance criterion) requires you to commensurately rescale the ε-balls in your Čech complex. Without such a corresponding rescaling on the output side, the scale invariance criterion cannot be coherently stated. The no-go result of 4.2.1 should thus neither be surprising nor worrisome.

4.3.4 Shoe-horning

Most practitioners will admit that the interpretation of homology in data is unclear. While increasing in popularity of late, TDA is still relatively niche. It is often reserved for situations in which traditional data analysis tools have failed to bear fruit, and TDA is one of many attempts to gain insight into the data—its more of a trial and error situation.

Since persistent homology has these nice properties, data scientists will often shoe-horn questions about data into the shape of a homology problem in order to make it tractable. For example, they might add extra edges to a Čech complex to turn open chains into closed loops. Or they might chose a particular dimensional reduction in which loops arise, as in Perea and Harer (2015). A fun example is the study of "tendrils", another geometric property of data clouds that is of potential interest. See image below—n tendrils emanating from a central cluster. By supplementing TDA with a procedure for identifying such clusters, these can be removed, and the tendrils can be tracked via the persistent H_0-homology of the resulting data cloud. Nicolau et al. (2011) use this technique to classify breast cancer types.

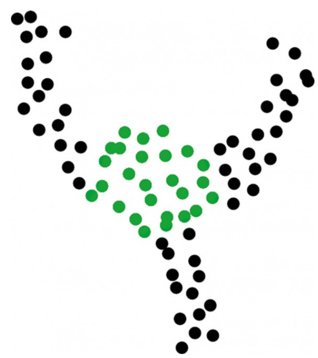


Figure 4.6: Visualization of data that features "tendrils", from Lesnick (2013).

Data scientists study persistent homology, not because they think of "counting holes" as the right way to characterize data, but rather because it is has really desirable features summarized by its functoriality. While the recent proliferation of these methods might be dismissed as mere hammer-nailing, it should rather be said that since we have very few tools to work with, we had better hope this problem can become nail-shaped.

4.3.5 Generalizing core concepts

Category theory is probably most "famous" for working at such a general, high level fashion as to draw analogies between seemingly disparate areas of mathematics. Bubenik and Scott (2014) show that the features of persistent homology that make it particularly conducive to data analysis can be summarized by the fact that its basic constructions constitute a special case of the much more general phenomenon.

Their paper suggests that barcodes can be used to generate meaningful persistence diagrams of features of data-clouds that are not merely topological, so long as the desired features satisfy some very general properties. Crucially, Bubenik and Scott prove a generalized version of the stability theorem of Cohen-Steiner et al. (2007). The abstract generality of category theory enabled Bubenik and Scott to point towards a much broader class of data analysis techniques that share the desirable properties of persistent homology. Again here, the category theory is not itself providing a tool for data analysis. Rather, category theory provides an abstract framework which "clarifie[s] the key ideas and proofs" and "allow[s] previous results to be vastly generalized" (Bubenik and Scott, 2014, p. 601).

4.4 TDA and spatial inference

4.4.1 Geometric understanding

Whether you think it is good or bad, it is an observational fact that visual, spatial, and aesthetic intuitions play a role in science. While it appears in more subtle ways, topological data analysts explicitly embrace the role of visual intuitions. As mentioned in section 4.3.4, TDA is a second-line resource for data that is particularly intractable to analyze, which puts creativity at the center of its application.

The goal of data analysis is to identify patterns in data that provide concise, comprehensible summaries of the system that point towards features of significance in broad classes of systems. Such recognition of patterns of sufficient generality without overfitting is the holy grail of artificial intelligence and machine learning research. In the mean-time, scientists still rely heavily on the *je ne sais quoi* features acquired through visual intuition to guide inquiry. To aid the evocations of these intuitions, data scientists will play around with parameters and data filtering. Since spatial intuitions exist at lower dimensions, the ability to use persistence modules to reduce dimensionality without losing information makes it especially useful.

While subjective visual judgments clearly dominate the earlier stages of inquiry, data analysts still return to more traditional empirical methods for post hoc justification. Even if a topological feature is robust under TDA analysis, the real measure of a successful analysis is whether it corresponds to a feature of the system of independent interest to scientists. Patterns found through random applications of TDA might lead scientists to look for such an independently interesting feature of a system, but if one cannot be found, the shapes identified in the data remain merely curiosities. In example 4.3, if barcodes did not track gender and age but some other feature that we do not independently classify as a natural kind, researchers would likely not have identified it. Even if they had stumbled upon a barcode pattern by chance, it would not have mattered if they could not tell a compelling story about what characteristic the pattern characterizes.

So spatial intuitions play a central role in the context of discovery, while their influence is fortified by the introduction of external empirical considerations at the stage of justification. But they reappear when the results are communicated to others, in the visual summary provided by a homological barcode or diagram. This allows data scientists to again invoke visual intuitions in evaluating the results of the analysis, which now contain all of the information about the persistence of shapes in an easily consumable, two-dimensional aid. The functoriality of TDA carries the visual information in CW-complexes through various reformulations until it finally reappears again in yet another visual format in its presentation.

4.4.2 Diagrammatic reasoning

Returning from our detour into the cognitive realm, we might wonder how all of this can be incorporated into a formal epistemic story about the structure of topological data models. Here, we can learn much from the vast literature on diagrammatic reasoning in Euclidean geometry. Critics of the rigor of reasoning from diagrams in geometric 'proofs' point to the fact that such proofs use a particular illustration to make an inference about all possible illustrations. However, philosophers of mathematical practice have recently come to appreciate the role of diagrams in generating and communicating geometric knowledge. Manders (2008) argues that ancient geometers were careful to rely on diagrams only for demonstrations about what he calls *co-exact* features—those that are relatively insensitive to the range of variation in possible visual representations, such as part-whole and boundary-interior relationships (and of course, homology). Mumma (2010) takes this a step further and develops a formal account of Euclidean proofs that includes both sentential and diagrammatic components.

How does this bear on TDA? Earlier, I noted that data analysts are concerned with ensuring that inferences about data rely only on real structural features of observations, rather than incidental features of how data is embedded in a metric space. At issue is the level of generality one can adopt when making inferences from a single visual representation of data, picked somewhat arbitrarily from an ensemble of possible alternative, equally valid representations. TDA resolves this issue by requiring that the analyzed features of data models be *functorial* with respect to maps that preserve what they take to be the relevant structural features of models, and *persistent* across parameters when the "right" value is not known.

4.4.3 Structure

The forgoing discussion about TDA hints at a new way to understand the relationship between the following two conceptions of "structure" in models of scientific theories:

1. The relevant causal and explanatory features of a system, abstracted from the noise present in any observation of the system; and
2. The content of a description of a physical system, abstracted from the particular language and formalism used to present it.

Data science largely concerns structure$_1$, while the philosophers of physics we met earlier are clearly concerned with structure$_2$. The relationship between these two notions at first appears relatively superficial. Yes, they are both ways of getting at what is *really there* in a physical system, but they seem to refer to completely different stages of scientific representation. The first comes in at the stage of observation and experimentation, referring to the structure of a particular physical system under observation. The second relates to extracting information from an idealized model (perhaps constructed out of "cleaned" data from the previous stage). The structure here consists of the components of the model that are actually doing the representational work, rather than merely scaffolding this content in language and symbols.

Scientific practice is a holistic process, and measurement cannot and should not be wholly separated from formal representation. Among other connections, observations inform how systems should be represented, and representations indicate directions for further research and measurement. Of course, no one would say these two senses of structure are mere homonyms, as they are clearly relying on similar intuitions about how structure is supposed go beyond particulars of an instance to something more essential. However, they seem to operate in different realms—structure$_1$ in the physical world and structure$_2$ in Platonic heaven—and thus they surely must be orthogonal to one another.

Nonetheless, I think TDA indicates how these two notions of structure are much closer than one might initially suspect. The first clue to this comes upon noticing that more than being intertwined with one another, particular acts of structural refinement by scientists may exist between realms. For example, suppose I collect demographic data that includes the hair color of participants, and include hair color as a feature of my initial abstract representation of this population, recorded as an RGB hex code. I then decide that precise hair color is not a relevant consideration for the theoretical purpose at hand, so I switch to presenting hair color information more coarsely as either light, medium, or dark. I can think of this "throwing away", "rounding off" or "smoothing out" as an act of cleaning data, obscuring noise at the observation level, and perhaps fundamentally changing the type of data I collect. Alternatively, I can think of it as refining my model—obscuring noise at the representational level. It amounts to the same act, viewed through different lenses. The moral is that the boundary between data and formalism is not completely clear.

Both definitions of structure invoke notions of robustness that are present in TDA. The functoriality of homology on the groupoid ensures our analysis respects structure$_1$—homology does not change when we switch between formally isomorphic representations. The functoriality of homology on the full category **CW** ensures it respects structure$_2$—it allows us to see how sensitive our analysis is to variations in parameter values like ε, where the relationships may not be isomorphisms.

4.5 Discussion

Data scientists are already aware of the fact that functoriality of homology is critical to TDA's utility in revealing and interpreting structural features of data sets. This paper offers an account of how and why this is this case, building on the analysis of TCM presented in previous chapters. There are various reasons to suspect that topological features correspond to meaningful signals in a data set. Moreover, topological features are accessible to visual cognition to aid in scientific interpretation. Since homology is *functorial* relative to the category (**CW**) that delineates the relevant structures, it is ensured that the reasons we had for thinking topological features were meaningful are preserved in the translation from data cloud to homological barcode.

Requiring functoriality constrains the tools that are available to us to analyze data, and homology is particularly well understood mathematically. Data scientists thus often try to apply persistent homology even if it is not immediately obvious why topological features of the data should be important. But by identifying that functoriality is operative in enabling robust inferences in TDA, we can use category theoretic tools to express the general features of *any* data analysis that might be epistemically sufficient. Bubenik and Scott (2014) provide the mathematical tools, and this paper supplements them by demonstrating how to construct an inferential narrative to justify their epistemic value. An obvious next step would be to explore new functorial data analysis methods (or functorializing old methods).

In the other direction, this discussion of TDA contributes to and clarifies the philosophical TCM project. Philosophers of science have spilt quite a bit of ink on how categories relate to representation in science. It is clear that they do, and in particular cases (mostly in classical physics) we can explain how and why, but the overall "theories as categories of models" program is far from well-developed. This exploration of TDA provides fresh perspective on TCM by considering it in a vastly different scientific context with different epistemic aims. There are four main takeaways here.

1. **Non-isomorphism relations matter.** Philosophers have focused on how TCM can elucidate equivalence between theories and models, and so have focused on groupoids rather than full categories. Morphisms that are not isomorphisms are necessary in

TDA, and including them reveals important information about part-whole relations between models.

2. **Formal and informal notions of structure are intimately connected.** Existing TCM literature employs a relatively abstract notion of structure. Section 4.4.3 reveals how TDA can provide a bridge between structure as intuitively meaningful features, and structure as formally meaningful mathematical constructions.

3. **Category theory ties geometric models to algebraic characterizations.** We already knew this because that's how category theory started (Eilenberg and MacLane, 1945), but one may have thought that was a historical accident. What previous applications in physics shared this character, philosophers have focused more on the abstract notion of intertranslatability between modeling schema than peculiarities of the relationships between algebraic and geometric theories. This feature is so important in TDA as to be difficult to ignore.

4. **Modeling schema are not fully reducible to category theory.** This relates to the previous point. The functoriality of homology endows homological barcodes with the structural content of persistence modules, but this relationship is not symmetric. Geometric models have special status in that they evoke knowledge generation through visual cognition. Category theory helps us externalize some of these pattern-recognition devices, but it certainly does not get us all the way there.

Taking these lessons to heart should inspire a richer understanding of how category theory bears on representation in science, and help guide future expositions and applications of TCM. And by considering TDA and TCM in tandem, we can learn a lot about how mathematical models represent structural features of the world.

Chapter 5

Conclusion

This text began in chapter 1 by presenting the "theories as categories of models" (TCM) framework for analyzing formalisms used to model physical systems. The proceeding chapters moved this discussion from abstraction to practical application, demonstrating how TCM can enrich scientific understanding. The question lingers, however, as to what role TCM itself performs in these applications.

At one extreme, there is a temptation to view TCM as a definitive method for establishing the content of and relationships between theoretical formalisms. While I doubt that many philosophers adopt this extreme view, it is natural to conceive of the case for TCM as a defense of its superiority relative to other methods of establishing theoretical (in)equivalence. I think this is misguided, but so is the other extreme view that takes TCM to be merely vacuous window-dressing for other, more fundamental accounts of formal structure.

My main takeaway from working with TCM is that it is an organizing principle for other accounts of scientific representation, but an extremely powerful one. TCM provides a scaffold for telling a story about how formalisms represent systems, and how they relate to one another, that is constrained to be coherent and defeasible. The process of defining a category to associate to a formalism amounts to being clear and precise about how the formalism represents physical systems. Defining a functor enforces similar clarity and precision in presenting the relationships between different physical formalisms. Presenting narratives in this way is not merely decorative—it exposes them to illuminating analysis using the tools of category theory. This text focused on one such tool—the property-structure-stuff heuristic—which elegantly summarizes the story captured by a functor. I suspect this only scratches the surface of what TCM has to offer philosophers of science.

Bibliography

Baez, J., T. Bartells, J. Dolan, and D. Corfield (2013). Property, Structure, and Stuff. `http://math.ucr.edu/home/baez/qg-spring2004/discussion.html`.

Barrett, J. W. (1991). Holonomy and path structures in general relativity and Yang-Mills theory. *International Journal of Theoretical Physics 30*(9), 1171–1215.

Barrett, T. W. (2014). On the Structure of Classical Mechanics. *The British Journal for the Philosophy of Science 66*(4), 801–828.

Barrett, T. W. (2015). Spacetime Structure. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics 51*, 37–43.

Barrett, T. W. (2018a). Equivalent and Inequivalent Formulations of Classical Mechanics. *The British Journal for the Philosophy of Science*.

Barrett, T. W. (2018b). What Do Symmetries Tell Us About Structure? *Philosophy of Science 85*(4), 617–639.

Barrett, T. W. and H. Halvorson (2016). Morita Equivalence. *The Review of Symbolic Logic 9*(3), 556–582.

Bednarczyk, M. A., A. M. Borzyszkowski, and W. Pawlowski (1999). Generalized Congruences— Epimorphisms in Cat. *Theory and Applications of Categories 5*(11), 266–280.

Bendich, P., J. S. Marron, E. Miller, A. Pieloch, and S. Skwerer (2016). Persistent Homology Analysis of Brain Artery Trees. *The Annals of Applied Statistics 10*(1), 198–218.

Bradley, C. and J. O. Weatherall (2019). On Representational Redundancy, Surplus Structure, and the Hole Argument. *arXiv preprint arXiv:1904.04439*.

Brocker, T. and T. tom Dieck (1985). *Representations of Compact Lie Groups*, Volume 98 of *Graduate Texts in Mathematics*. Berlin, Heidelberg: Springer Berlin Heidelberg.

Bubenik, P. (2015). Statistical Topological Data Analysis Using Persistence Landscapes. *Journal of Machine Learning Research 16*, 77–102.

Bubenik, P. and J. A. Scott (2014). Categorification of Persistent Homology. *Discrete & Computational Geometry 51*(3), 600–627.

Caetano, A. and R. F. Picken (1994). An Axiomatic Definition of Holonomy. *International Journal of Mathematics 05*(06), 835–848.

Carlsson, G. (2009). Topology and Data. *Bulletin of the American Mathematical Society 46*(2), 255–308.

Carlsson, G., T. Ishkhanov, V. De Silva, and A. Zomorodian (2008). On the Local Behavior of Spaces of Natural Images. *International Journal of Computer Vision 76*(1), 1–12.

Carlsson, G. and F. Mémoli (2008). Persistent Clustering and a Theorem of J. Kleinberg. *arXiv preprint arXiv:0808.2241*.

Cohen-Steiner, D., H. Edelsbrunner, and J. Harer (2007). Stability of Persistence Diagrams. *Discrete Comput. Geom. 37*(1), 103–120.

Crawley-Boevey, W. (2015). Decomposition of Pointwise Finite-Dimensional Persistence Modules. *J. Algebr. Appl. 14*(05), 1550066.

Doran, C., A. Lasenby, and J. Lasenby (2003). *Geometric Algebra for Physicists*. Cambridge University Press.

Earman, J. (1986). Why Space is not a Substance (at least not to first degree). *Pacific Philosophical Quarterly 67*(4), 225–244.

Edelsbrunner, H., D. Letscher, and A. Zomorodian (2002). Topological Persistence and Simplification. *Discrete & Computational Geometry 28*, 511–533.

Efron, B. (1979). Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics 7*(1), 1–26.

Eilenberg, S. and S. MacLane (1945). General Theory of Natural Equivalences. *Transactions of the American Mathematical Society 58*(2), 231–294.

Gambini, R. and J. Pullin (2000). *Loops, Knots, Gauge Theories and Quantum Gravity*. Cambridge University Press.

Geroch, R. (1972). Einstein Algebras. *Communications in Mathematical Physics 26*(4), 271–275.

Ghrist, R. (2008). Barcodes: The Persistent Topology of Data. *Bulletin of the American Mathematical Society 45*(1), 61–75.

Ghrist, R. (2014). *Elementary Applied Topology* (1 edition ed.). s. l.: CreateSpace Independent Publishing Platform.

Giles, R. (1981). Reconstruction of Gauge Potentials from Wilson Loops. *Physical Review D 24*(8), 2160–2168.

Halvorson, H. (2012). What Scientific Theories Could Not Be. *Philosophy of Science 79*(2), 183–206.

Halvorson, H. and D. Tsementzis (2015). Categories of Scientific Theories. In E. Landry (Ed.), *Categories for the Working Philosopher*, pp. 402–429. OUP.

Hatcher, A. (2002). *Algebraic Topology*. Cambridge UP, Cambridge.

Healey, R. (2007). *Gauging What's Real*. Oxford University Press.

Heller, M. (1992). Einstein Algebras and General Relativity. *International journal of theoretical physics 31*(2), 277–288.

Hestenes, D. and G. Sobczyk (2012). *Clifford Algebra to Geometric Calculus: a Unified Language for Mathematics and Physics*, Volume 5. Springer Science & Business Media.

Hudetz, L. (2019). Definable Categorical Equivalence. *Philosophy of Science 86*(1), 47–75.

Huggett, N. and C. Hoefer (2018). Absolute and Relational Theories of Space and Motion. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2018 ed.). Metaphysics Research Lab, Stanford University.

Kleinberg, J. M. (2003). An Impossibility Theorem for Clustering. In *Advances in Neural Information Processing Systems*, pp. 463–470.

Kobayashi, S. and K. Nomizu (1996). *Foundations of Differential Geometry, Vol.1*. New York: Wiley-Interscience.

Lee, A. B., K. S. Pedersen, and D. Mumford (2003). The Nonlinear Statistics of High-Contrast Patches in Natural Images. *International Journal of Computer Vision 54*(1-3), 83–103.

Lesnick, M. (2013). Studying the Shape of Data Using Topology. `https://www.ias.edu/ideas/2013/lesnick-topological-data-analysis`. Accessed: 2018-11-19.

Mac Lane, S. (2013). *Categories for the Working Mathematician*, Volume 5. Springer Science & Business Media.

Malament, D. B. (2012). *Topics in the foundations of general relativity and Newtonian gravitation theory*. University of Chicago Press.

Manders, K. (2008). The Euclidean Diagram (1995). In *The Philosophy of Mathematical Practice*. Oxford: Oxford University Press.

Mumma, J. (2010). Proofs, Pictures, and Euclid. *Synthese 175*(2), 255–287.

Nestruev, J. (2006). *Smooth Manifolds and Observables*, Volume 220. Springer Science & Business Media.

Nguyen, J., N. J. Teh, and L. Wells (2017). Why Surplus Structure is not Superfluous. *The British Journal for the Philosophy of Science*.

Nicolau, M., A. J. Levine, and G. Carlsson (2011). Topology Based Data Analysis Identifies a Subgroup of Breast Cancers with a Unique Mutational Profile and Excellent Survival. *Proceedings of the National Academy of Sciences 108*(17), 7265–7270.

Perea, J. A. and J. Harer (2015). Sliding Windows and Persistence: An Application of Topological Methods to Signal Analysis. *Foundations of Computational Mathematics 15*(3), 799–838.

Rejzner, K. (2016). *Perturbative Algebraic Quantum Field Theory*. Mathematical Physics Studies. Cham: Springer International Publishing.

Rosenstock, S., T. W. Barrett, and J. O. Weatherall (2015). On Einstein Algebras and Relativistic Spacetimes. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics 52, Part B*, 309–316.

Rosenstock, S. and J. O. Weatherall (2016). A Categorical Equivalence between Generalized Holonomy Maps on a Connected Manifold and Principal Connections on Bundles over that Manifold. *Journal of Mathematical Physics 57*(10), 102902.

Rynasiewicz, R. (1992). Rings, Holes and Substantivalism: On the Program of Leibniz Algebras. *Philosophy of Science 59*(4), 572–589.

Schreiber, U. and K. Waldorf (2009). Parallel Transport and Functors. *J. Homotopy Relat. Struct. 4*, 187–244. arXiv: 0705.0452.

Sober, E. (2015). *Ockham's Razors*. Cambridge University Press.

van de Weygaert, R., G. Vegter, H. Edelsbrunner, B. J. T. Jones, P. Pranav, C. Park, W. A. Hellwing, B. Eldering, N. Kruithof, E. G. P. p. Bos, J. Hidding, J. Feldbrugge, E. ten Have, M. van Engelen, M. Caroli, and M. Teillaud (2011). Alpha, Betti and the Megaparsec Universe: On the Topology of the Cosmic Web. In M. L. Gavrilova, C. J. K. Tan, and M. A. Mostafavi (Eds.), *Transactions on Computational Science XIV: Special Issue on Voronoi Diagrams and Delaunay Triangulation*, pp. 60–101. Berlin, Heidelberg: Springer Berlin Heidelberg.

Weatherall, J. O. (2015). Fiber bundles, Yang-Mills theory, and general relativity. *Synthese 193*(8), 2389–2425.

Weatherall, J. O. (2016a). Are Newtonian Gravitation and Geometrized Newtonian Gravitation Theoretically Equivalent? *Erkenntnis 81*(5), 1073–1091.

Weatherall, J. O. (2016b). Regarding the Hole Argument. *The British Journal for the Philosophy of Science 69*(2), 329–350.

Weatherall, J. O. (2016c). Understanding Gauge. *Philosophy of Science 83*(5), 1039–1049.

Weatherall, J. O. (2017). Categories and the Foundations of Classical Field Theories. In E. Landry (Ed.), *Categories for the Working Philosopher*, pp. 329–348. Oxford University Press.

Wu, T. T. and C. N. Yang (1975). Concept of nonintegrable phase factors and global formulation of gauge fields. *Physical Review D 12*(12), 3845–3857.

Zomorodian, Afra and Carlsson, Gunnar (2008). Localized homology. *Computational Geometry 41*(3), 126–148.