UC San Diego
UC San Diego Electronic Theses and Dissertations

Title
Identity in Gauge Theories

Permalink
https://escholarship.org/uc/item/91h16091

Author
Dougherty, John

Publication Date
2018

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA SAN DIEGO

Identity in Gauge Theories

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Philosophy

by

John Dougherty

Committee in charge:

Professor Craig Callender, Chair
Professor Nancy Cartwright
Professor David Meyer
Professor Charles Sebens
Professor James Owen Weatherall

2018

Copyright
John Dougherty, 2018
All rights reserved.

The dissertation of John Dougherty is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California San Diego

2018

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGEMENTS

This work was performed in part under a collaborative agreement between the University of Illinois at Chicago and the University of Geneva and made possible by grant (10) 56314 from the John Templeton Foundation, and its contents are solely the responsibility of the author and do not necessarily represent the official views of the John Templeton Foundation.

Chapter 2, in full, is a reprint of material currently being prepared for submission for publication. The dissertation author is the sole author of this material.

Chapter 3, in full, is a reprint of the material as it appears in Dougherty, J. (2017). "Sameness and separability in gauge theories", *Philosophy of Science*, 84(5), 2017. The dissertation author was the sole author of this paper.

Chapter 4, in full, is a reprint of material currently being prepared for submission for publication. The dissertation author is the sole author of this material.

I have benefited from the help of many people in the completion of this dissertation. I am grateful to Craig Callender, Nick Huggett, Jim Weatherall, and Chris Wüthrich for extensive professional supervision and support. I am grateful to Craig Agule, Joyce Havstad, Casey McCoy, Ben Sheredos, and Adam Streed for extensive feedback and mentorship. I am grateful *simpliciter* to Kathleen Connelly.

VITA

2010	A. B. in Physics with Honors, University of Chicago
2010	A. B. in the History, Philosophy, and Social Studies of Science and Medicine with Honors, University of Chicago
2015	C. Phil. in Philosophy, University of California San Diego
2018	Ph.D. in Philosophy, University of California San Diego

PUBLICATIONS

Dougherty, J. and Callender, C. (Forthcoming). Black hole thermodynamics: More than an analogy? In Ijjas, A. and Loewer, B., editors, *Guide to the Philosophy of Cosmology*. Oxford University Press.

Dougherty, J. (Forthcoming). What inductive explanations could not be. *Synthese*.

Dougherty, J. (2017). Sameness and separability in gauge theories. *Philosophy of Science*, 84(5):1189–1201.

ABSTRACT OF THE DISSERTATION

Identity in Gauge Theories

by

John Dougherty

Doctor of Philosophy in Philosophy

University of California San Diego, 2018

Professor Craig Callender, Chair

I argue that contemporary theories of physics are best understood as involving a more refined identity relation than the one found in first order logic. I use this identity relation to provide novel interpretations of Yang–Mills theory and General Relativity, two of the most successful physical theories. I show that my interpretation is superior to those currently on the market because it solves problems left open by extant interpretations and provides a better explanation of the distinctive features of these theories. My interpretation also fills gaps in justification found in the physics literature.

1 Introduction

In this dissertation I give an account of symmetry and its relationship to identity, and I apply this account to the interpretation of physical theories.

I am primarily concerned with the ways that symmetry is used as a guide to the content of a theory. This is not symmetry's only philosophical dimension; symmetry considerations are ubiquitous in physics and its philosophy, and play many distinct (though related) roles. For example, first-order debates in the philosophy of physics are sometimes generated by the symmetries of particular theories. Arguments over the existence and nature of space and time appeal to symmetries of translation, rotation, and reflection. The permutation symmetries of quantum mechanics complicate the concept of individuality. And modern theories of particle physics feature symmetries whose importance is matched only by their obscurity. Symmetries also impinge on more general philosophical debates over theoretical equivalence, the progress of science, the empirical content of a theory, and more.[1]

The story I tell here has consequences for these debates, but my focus will be on arguments that use symmetries to suss out a physical theory's commitments. These arguments are all guided by the following maxim: a theory countenances some feature as

[1] On spacetime arguments, see Earman (1989) and Pooley (2013); on quantum mechanics French and Redhead (1988), Huggett (1999), and Caulton (2013). On the Yang–Mills theories of modern particle physics, see Belot (1998), Healey (2007), Weatherall (2016b) and below. The other topics I list are addressed by Black (1952), Post (1971), Hacking (1975), Redhead (1975), Kosso (2000), Belot (2001), Roberts (2008), Greaves and Wallace (2014), Caulton (2015), Rosenstock et al. (2015), and my discussion below.

real only if that feature is invariant with respect to the symmetries of the theory. Thus Leibniz argues that space has no independent existence due to spatial translation symmetry, and Einstein argues against absolute simultaneity from the symmetries of electrodynamics. "If a putative feature is variant in laws that we have reason to think are true and complete, then this is some reason to think that the feature is not real" (2018, 840). The philosophical and physical literatures are littered with cases like this. Many philosophers have offered analyses of this argument form along Occamist lines.[2] These accounts are usually offered as analyses of a widely-accepted pattern of inference—that is, they seek to explain why we might obey this maxim and why it has been successful, rather than defend it against detractors.[3] And they take this maxim to be universally applicable. If some feature is variant, this is always a reason to think it is not real. Borrowing terminology from Shamik Dasgupta (2018), call this approach to physical interpretation the "method of symmetry".

The method of symmetry seems to be clearly justified in some cases, and it's getting more popular to think that it's always justified.[4] It think this popularity is well-deserved: the method of symmetry *is* always justified. But now suppose we find some feature that is variant under some symmetry—the absolute velocity of the solar system, say, or the rest frame of the luminiferous ether—and we conclude that it is not real. The question motivating this dissertation is: now what? As I argue in this chapter, important options at this stage have been neglected. Of course, if some theory includes some superfluous feature then the natural prescription is removal of that feature, producing a more accurate theory.

[2] Borrowing Dasgupta's (2016) taxonomy, we might take Earman (1989), van Fraassen (1989), North (2009), Baker (2010), and Belot (2013) to be offering readings in terms of redundancy, Weyl (1952) and Nozick (2001) to be offering readings in terms of objectivity, and Ismael and van Fraassen (2003), Roberts (2008), and Dasgupta (2016) to be offering readings in terms of epistemic considerations. Not all of these authors take these inferences to be justified (at least not in all forms). I am most sympathetic to the redundancy analysis, at least insofar as it can be separated from the epistemic analysis.

[3] These detractors include Belot (2013) and Møller-Nielsen (2017).

[4] As Belot (2018, fn. 27) says, "some see something like an emerging consensus among philosophers of physics" that the method of symmetry is universally applicable. "Baker (2010, 1157) speaks of this view as being 'widely held,' Greaves and Wallace (2014, 60) speak of 'a widespread consensus around it, Teh (2016, 98) of it as agreed upon by many not all" (Belot, 2018, 29). Metaphysicians, as Belot notes, are split on the question.

But the details of this surgery matter, and it isn't always carried out successfully. Problems then trickle up: the inadequacies of the resulting theory can be used as a *modus tollens* argument against the maxim of the previous paragraph, and this has knock-on effects for the other uses of symmetries in philosophy of physics.[5] In this chapter I discuss the problem in general and indicate an alternative technique for excising superfluous structure from a mathematized physical theory. We can therefore retain the method of symmetry in the face of apparent counterexamples, and we get more besides. In the following chapters I use this technique to address particular puzzles about particular theories.

The word "symmetry" is used variously as a one-place predicate, a two-place predicate, and a count noun. The next section discusses each of these uses in some detail. I argue that we should treat numerical identity analogously. Any object has the property of "being self-identical", and this means that the object bears the identity relation to itself. I claim that, furthermore, "identity" should sometimes be used as a count noun, such that an object can bear an identity to itself—and, possibly, more than one. Perhaps this way of speaking is obscure; I explain in detail below. To lay the groundwork for this explanation, I first argue that symmetry comes to play its guiding role through its relationship to underdetermination; symmetries are "beacons of redundancy" (Ismael and van Fraassen, 2003, 391). This explains why the usual approach to removing superfluous structure—quotienting—is a solution to the problems posed by symmetry: it removes underdetermination. But if we're not careful then quotienting can remove too much. I show that adding identities (in the count noun sense) gives a more conservative alternative to the standard approach to quotienting as a way to remove superfluous structure and avoid underdetermination.

My goal in introducing count-noun identities and modifying the standard quotient procedure is to give an account of the difference between gauge symmetries and physical

[5] Belot (2007, 2013, 2018) in particular has made a number of arguments along these lines.

symmetries. The notion of a count-noun identity is much more general than this, and this chapter is primarily concerned with simple examples that don't presuppose familiarity with the most important examples of theories that feature gauge symmetries.[6] But my discussion in this chapter is ultimately in the service of giving an account of the difference between gauge symmetries and other symmetries in these theories (viz., Yang–Mills theory and general relativity). There is little agreement about what a gauge symmetry is, and therefore little agreement about which symmetries classify as "gauge". But by and large there *is* agreement that they indicate redundancies in the mathematical representation. That is, they are paradigmatic targets for the method of symmetry. However, there have as yet been no satisfactory applications of the method of symmetry to these theories. As Redhead (2003) explains, we run into problems whether we eliminate symmetry-variant features or leave them in. More recently, Healey (2007) has argued that we can, in fact, apply the method of symmetry in Yang–Mills theory, if we are sufficiently careful. But as I argue below, Healey (2007)'s interpretation is inadequate.

So, in brief, in this dissertation I argue that gauge symmetries should be interpreted as count-noun identities, which non-gauge symmetries should be interpreted as relating distinct physical states of affairs. And I claim that this alternative gives the best account of certain features of Yang–Mills theory and general relativity. The following chapters consider some examples of this.

[6] More generally, I take count-noun identities to be the subject of homotopy theory. Like Barwick (2017), I take homotopy theory to be "an enrichment of the notion of equality, dedicated to the primacy of *structure over properties*" (2017, 2). This statement is made precise in homotopy type theory, for which see Univalent Foundations Program (2013) for a textbook treatment or Ladyman and Presnell (2017) for a philosophical introduction.

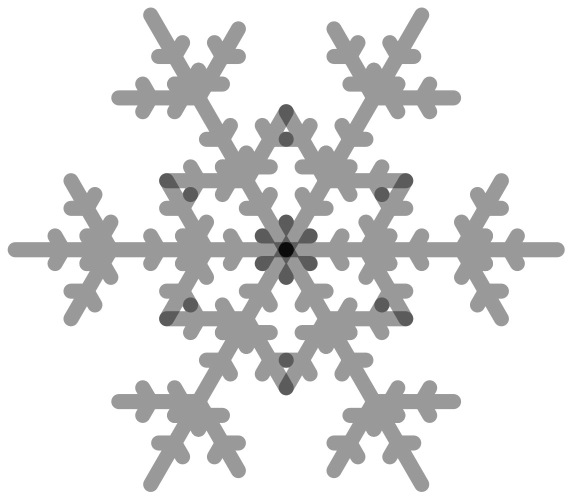

Figure 1.1: A snowflake

1.1 Symmetry and underdetermination

Symmetries serve as a guide to a theory's commitments because they have a systematic relationship to underdetermination and we have a systematic way of exterminating them. Symmetries always lead to underdetermination of some kind. And from Occamist considerations we have reason to minimize underdetermination. As we will see in the next section, symmetries are easy to eliminate; this means that the underdetermination they lead to is also easy to eliminate.[7] Thus symmetries are a guide to which parts of a theory might be easily eliminable. The connection between symmetry and underdetermination comes about as follows.

Start with "symmetry" in the property sense, and consider again a snowflake with six-fold rotational symmetry, such as in Fig. 1.1. This symmetry is a regularity in the relationships between the snowflake's parts, and it prevents us from giving a definite description of every part of the snowflake. The point at the center of the snowflake is an exception, of course—"the point at the center" uniquely refers. But for any other point on the snowflake and any qualitative description of that point in terms of the other parts of

[7] How does eliminating symmetry relate to eliminating the things that are variant under the symmetry

the snowflake there are at least five other points satisfying the same description. There are six outermost tips, six tips on the inner star, and so on. So in general the referent of qualitative descriptions involving only parts of the snowflake will be underdetermined. Other descriptions can overcome this underdetermination—I can refer to the outermost point on the right side of the page, for example—but any symmetry in this sense undermines the referential ability of a corresponding set of descriptive resources.

The philosophically serious version of this example concerns the reference of theoretical terms. According to the tradition of Frank Ramsey (1931), Rudolf Carnap (1956, 1966), David Lewis (1970), and Adam Caulton (2015), we refer to unobservable entities and kinds by definite description: a theoretical term refers to the unique thing that plays the theoretical role of that term.[8] But if the world is symmetric with respect to the descriptive resources of the theory then there won't be a unique thing that plays a given theoretical role, just as there isn't a unique point on the snowflake answering to most qualitative descriptions involving other parts of the snowflake. So in the presence of symmetries we cannot secure reference for our theoretical terms by this method. Indeed, Lewis (1970, 433) argues that in a case like this we should say that our theoretical term is denotationless. He later saw that this problem can sometimes be avoided in the same way we broke the underdetermination in the case of the snowflake: appeal to ostension and contingent features of our world. Positive and negative charge have symmetric roles in the laws of physics, but

> negative charge is found in the outlying parts of atoms hereabouts, and positive

[8] This is only a rough-and-ready characterization of this tradition, and it papers over important differences. While Ramsey and Carnap are looking to replace all theoretical terms in one go, Lewis and Caulton mean to give the meaning of each new theoretical term as it is introduced. So, in particular, new theoretical terms can be defined using a description that includes existing theoretical terms. And it's not quite right to say that these proposals assign to a theoretical term the unique entity satisfying some description. For example, Lewis adopts Scott's (1967) account of denoting phrases—as opposed to Russell's (1905) analysis in terms of camouflaged existential quantification or Frege's (1892) or Strawson's (1950) gappy analyses—in order to correctly handle the cases in which the theoretical term does not refer. Since we're just trying to see why symmetries pose a problem for this kind of analysis these details aren't germane.

> charge is found in the central parts. ... Thus the theoretical roles of positive and negative charge are not purely nomological roles; they are locational roles as well (Lewis, 2009, 207).

But this move only works when some other descriptive resources can break the symmetry. In particular, we cannot appeal to it in the case of fundamental theories of physics, which are meant to provide complete descriptions of the state of the world.

Symmetry properties are accompanied by symmetry relations. These come in two kinds. The first kind holds between the parts of the object with the symmetry property. Two points on the snowflake are symmetric with respect to the structure of the snowflake if they satisfy exactly the same qualitative descriptions involving the parts of the snowflake. For any point p on the snowflake other than the center there are six points q such that p and q stand in this symmetry relation. And these are exactly the points that undermine the referential capacities of qualitative descriptions of the snowflake's parts. So the symmetry relations that hold between the parts of the snowflake are another expression of the fact that the snowflake has the property of rotational symmetry.

The second kind of symmetry relation can lead to empirical underdetermination. Suppose that you leave the room, I rotate the snowflake, and you return. If I rotated the snowflake 60° or 120° then it would look exactly the same as if I'd left it where it was. No matter what observations of the location facts you undertake, you will be unable to determine whether and how much I rotated the snowflake. So the configurations of the snowflake before and after you stepped out are symmetric with respect to the location facts. As before, appealing to other features might allow you to break the empirical underdetermination, but fixing all of the observable location facts leaves open six possible changes in configuration. There's some fact about how the extension has changed, but this fact is inaccessible to you through observation of location alone.

This example has a lot of moving parts, and it's worth disentangling them to distinguish the specifics of the example from the general connection between symmetry and underdetermination. In particular, I want to be clear that this example has nothing to do with the exact contours of (un)observability. The symmetry relation of the previous paragraph is synonymous with the indiscernibility relation induced by the observable features of the snowflake. More generally, for any collection of objects any class of features induces an indiscernibility relation on the collection of objects: two objects are indiscernible if they agree on all of the features in the class. So the locational features of the snowflake induce an indiscernibility relation on the configurations of the snowflake, and two configurations are symmetric with respect to these features if they bear this relation to one another. And mass, shape, and magnetism induce an indiscernibility relation on bits of metal, such that if a slug and a quarter are symmetric with respect to these features they have the same purchasing power at a vending machine. So the underdetermination in the case of the snowflake doesn't arise from ambiguities in the split between observable and unobservable features of the snowflake. Rather, any class of features induces an indiscernibility relation, and if any distinct objects bear this relation to one another then checking these features won't suffice to distinguish them. A vending machine will either accept what you put in or it won't. If it will, then what you've deposited is symmetric with legal tender with respect to these features—and if there isn't much more symmetry then you've an ill-gotten snack headed your way.

Philosophical uses of symmetry in the relational sense usually involve a negotiation between the amount of underdetermination, the class of relevant features, and the collection of things bearing these features. Since underdetermination arises from too much symmetry, it can be avoided by eliminating symmetries. As an example, consider Pierre Duhem's (1914, VI.2) argument against the possibility of an experiment that tests a single hypothesis. To test some hypothesis a physicist deduces an observable consequence of this hypothesis and heads to the lab to look for this consequence. But as Duhem (1914) points out, this deduction involves not just the hypothesis up for debate but also "the whole theoretical scaffolding used by the physicist" (1962, 185). If observations are at odds with the prediction and the deduction is valid, then one of the premises must be false. But in a valid deductive argument the premises make symmetric contributions to the conclusion: if the conclusion is false one of the premises must be false as well, but it is underdetermined which one. So an observation at odds with the prediction cannot tell us whether we ought to reject the hypothesis at issue or some part of the theoretical scaffolding.

We can understand responses to this underdetermination as seeking to reduce the symmetry of the premises so as to make one premise the one we ought to reject.[9] Duhem (1914, VI.10) argues that the physicist's "good sense" can break the symmetry and identify some premises as more likely candidates for rejection than others. More concretely, as Larry Laudan argues, "no serious twentieth-century methodologist has ever espoused, without crucial qualifications, logical compatibility with the evidence or logical derivability of the evidence as a sufficient condition for detachment of a theory" (1990, 278). So even if every premise is logically symmetric with respect to the conclusion we should expect some rational asymmetries. And Helen Longino (1990) and Elizabeth Anderson (1995), among others, have argued that there may be social reasons to reject one premise or another. Each of these responses looks to overcome the underdetermination by increasing the class of relevant features, making the symmetry relation stronger. Alternatively, one could shrink the collection of things bearing these features. This is the quotient method, which I discuss in the next section.

Consider, finally, "symmetry" as a count noun. The property and relation senses of "symmetry" are familiar: the former is visually impressive, the latter is synonymous with the familiar indiscernibility relation. But the sense of "symmetry" at issue in this dissertation is the count-noun sense, which is perhaps less familiar. In this sense a symmetry is usually

[9] In this paragraph I run together various different kinds of underdetermination discussed by Duhem (1914), Quine (1961), Sklar (1975), Stanford (2001), and others. Since these kinds of underdetermination don't play any role in my argument (aside from serving as examples), this equivocation is harmless. But see Turnbull (2018) for disambiguation.

characterized as "a transformation... which preserves certain salient features" (2015, 155), and it is the most common sense of "symmetry" in the philosophy of physics literature.[10] Returning to the snowflake for an example, recall that every configuration of a snowflake stands in a symmetry relation with the configuration that would be obtained by rotating the snowflake 60°. Rotating the snowflake by 60° is therefore an example of a symmetry transformation. This transformation sends the center point to itself, it sends the outermost tip on the right to the outermost tip in the upper right, and so on. Other symmetries of the snowflake include rotation by 120°, left–right reflection, and moving the snowflake one inch up then one inch down. Symmetry transformations differ from the symmetry relations discussed above because relations are extensional in a way that symmetry transformations aren't. A rotation by 360° takes every configuration to itself, but it isn't the same as the transformation that doesn't do anything.

This characterization of count-noun symmetries appeals to an intuitive notion of "transformation" that needs clarifying for our discussion below. On the one hand, there are many uncontroversial examples of count-noun symmetries that are not transformations in any usual sense. On the other, there are many transformations that I do not want to call count-noun symmetries. If the snowflake were nailed down then "rotation by 60°" would still be a symmetry transformation of the snowflake, though we couldn't rotate it. And a ball of uranium more than a mile in diameter would have rotational symmetry transformations, though there can't be such a ball. This disconnect is even clearer when we talk about possible worlds or situations related by a "transformation". Philosophers sometimes discuss "shifting" a possible world, even though a possible world can't be the direct object of some behavior. So if we talk about symmetry "transformations" of these,

[10] Similar characterizations are given by the contributions to Brading and Castellani (2003b), Baker (2010), Belot (2013), Dasgupta (2016), and most everyone else. Weatherall (2016b) and Barrett (2018) give slightly more sophisticated characterizations in terms of invertible arrows in a category of relevant structures. Their notion of symmetry is closer to what I intend by "count-noun symmetries", though I want to generalize further to equivalences in higher categories.

it must generally be in a metaphorical sense. That we are able (when conditions are right) to manipulate symmetric objects in ways that preserve some salient facts is a consequence of the existence of count-noun symmetries. The manipulations are not themselves the symmetry.

To avoid the tempting connotations of the word "transformation", I will continue to call symmetries in this sense "count-noun symmetries". Officially, a count-noun symmetry is an element of a symmetry group, and a symmetry group is a group acting on a set in a way that preserves some relevant properties, relations, or structures.[11] Since the usual recipe for symmetry elimination also involves these concepts, it's worth spelling out these concepts in mathematical detail.

Definition 1.1.1. A *group* G consists of

- a set G of *elements* of the group
- for any elements g and g' of G an element $g \cdot g'$ of G, the *product* of g and g', such that
- for all elements g, g', and g'' of G we have $(g \cdot g') \cdot g'' = g \cdot (g' \cdot g'')$,
- an element e of G, the *identity* element, which satisfies $g \cdot e = g = e \cdot g$ for all g in G, and
- for every element g of G an element g^{-1}, the *inverse* of g, such that $g^{-1} \cdot g = e = g \cdot g^{-1}$.

Definition 1.1.2. Let X be a set and G a group. An *action* of G on X consists of

- a map sending an element g of G and an element x of X to an element $g \cdot x$ of X such that

[11] This characterization of symmetry is ubiquitous; for example, see Brading and Castellani (2003a), Belot (2003), and Baker (2010, fn. 2).

- we have $e \cdot x = x$ for e the identity element of G
- for all g and g' in G we have $(g \cdot g') \cdot x = g \cdot (g' \cdot x)$.

To say that a group G acting on a set X is a symmetry group with respect to some property P means that $P(x)$ if and only if $P(g \cdot x)$ for all x in X and g in G. So, for example, let X be the set of possible configurations of our snowflake. For every integer n we have a mapping that sends any configuration of the snowflake to the configuration that results from rotating the snowflake n times about its center by 60°. Since this produces a configuration with the same observable spatial features the group of integers is a symmetry group, and there is one count-noun symmetry of the snowflake for every integer.

This characterization of count-noun symmetries avoids the problems with the characterization in terms of transformations. The action of the group on the set allows us to speak of "applying" symmetries, since any symmetry g can be thought of as a map that sends an element x of X to the element $g \cdot x$. But this kind of map is just a relation, so a symmetry has the desirable features of a transformation without the problematic connotations of activity. And it's compatible with the way we usually talk about symmetries in terms of physical manipulations. We can refer to a group element by the functional relation it induces via some action. For example, the count-noun symmetry referred to as "rotation by 60°" is the input–output relation of the physical activity of rotating some object by 60°. Since the same group can act on many other sets, we can say that any radially symmetric object has the count-noun symmetry, even if the physical activity of rotation can't be effected.

Where underdetermination is concerned, count-noun symmetries are traditionally thought to be relevant because of the symmetry relations they induce. Any class of count-noun symmetries on a collection generates an equivalence relation on that collection by taking two elements to be equivalent if they are related by some symmetry. In the snowflake case this induced relation is the symmetry relation considered above. Because rotation is a symmetry it preserves some salient feature, so the configurations it relates agree on this feature and we have the same kind of underdetermination as in the case of relations. And, as always, the existence of this relation means that the collection has a symmetry property, just as the collection of points making up the snowflake has rotational symmetry in virtue of the symmetry relations between them.

One of my main claims in this dissertation is that the relationship between count-noun symmetries and underdetermination is not entirely captured by the symmetry relation induced by the count-noun symmetry. Two different classes of count-noun symmetries can induce the same symmetry relation while leading to underdetermination of different kinds—and, therefore, underdetermination that should be fixed in different ways. For example, the six-element set of rotations of the snowflake by $60°$, $120°$, $180°$, $240°$, $300°$, and $360°$ induces the same symmetry relation on the points of the snowflake as the infinite set containing all integer multiples of $60°$. Despite inducing the same symmetry relation, the difference between these classes can have physical consequences—it appears, for example, in the analysis of the rotation of an electron or other particles with non-integer spin.

Since the connection between count-noun symmetries and underdetermination traditionally goes through the induced symmetry relation, extant examples of this connection in the philosophical literature are generally of the same character as the last example we considered. Thus philosophers argue that the translation symmetry of Newtonian mechanics means that we have no empirical access to an object's absolute position, and the Galilei boost symmetry of Newtonian mechanics means we have no empirical access to an object's absolute velocity. An important place where count-noun symmetries go beyond the relations they induce can be found in so-called "hole" arguments, which show that count noun symmetries result in predictive underdetermination in some theories. The most famous example of this kind of argument arises in general relativity, as we discuss in Chapter 4, though it has also been deployed in the case of Yang–Mills theories.[12] Hole arguments show that a count-noun symmetry can make the predictions of a theory radically underdetermined.

As an illustration, consider the following toy economic example.[13] Redenomination is, ideally, an economic symmetry. In 1960 France introduced the new franc, which was worth 100 francs. Over the next three years the old francs left circulation and "new" was dropped from the name of the new francs. Because the exchange rate between new and old francs was fixed, this exchange was an economic symmetry. If X was richer than Y in old francs, then X was richer than Y in new francs, 100 old francs had the exact same purchasing power as 1 new franc, and so on. So the economic state of affairs before and after the redenomination is the same. This means that the price of eggs is completely unpredictable by any economic theory. For suppose some theory says that a dozen eggs could cost \$3.00 in 2025. Since the theory is indifferent between economically symmetric states of affairs, it must also say that a redenomination is possible before then. And so it must also say that a dozen eggs could cost \$0.03 in 2025, since the dollar might be redenominated in 2020 in the same way the franc was in 1960. Indeed, *any* price must be possible, since the dollar could be arbitrarily redenominated. So the theory puts no constraints on the price of eggs.

So the three senses of "symmetry" just reviewed are systematically related to one another and to underdetermination. The relationship to underdetermination is what makes them such a good guide to the commitments of a theory, for reasons to be explained presently. But the interrelations of the three senses of symmetry can mislead. In particular, it's important to distinguish a collection of count-noun symmetries from the symmetry

[12] On the hole argument in general relativity, see Chapter 4 and the references there. For a version of the hole argument in Yang–Mills theories, see Lyre (1999).

[13] The analogy between foreign exchange markets and lattice gauge theories that underlies this illustration comes from Malaney (1996). Maldacena (2015) gives a very readable exposition of this analogy to help explain the Higgs mechanism.

relation they induce. If we fail to keep track of this distinction then we will end up removing too much when we apply the symmetry-elimination method I now describe.

1.2 Removing underdetermination

Caulton expresses a common sentiment when he says that "[t]he existence of non-trivial symmetries... prompts a *quotienting* of our original space of mathematical states" (2015, 156). I take this to be one formalization of the guiding maxim from the opening: a theory only countenances features that are invariant under the symmetries of the theory. And I take this, in turn, to follow from Occamist reasoning. The connection between these three, on my view, is due to the relationship between symmetry and underdetermination. We should only admit underdetermination into our theories if it has some payoff. So we should only admit symmetries if the underdetermination they engender has some payoff. Quotienting is a way to remove symmetry, and so we are led to Caulton's claim. But there are many kinds of quotienting, and philosophers and physicists have erred in focusing on the most familiar one—corresponding to the relational sense of symmetry—and neglecting the kind of quotient applicable to count-noun symmetries.

The force that's required of the Occamist norm in this reasoning will vary from case to case. A general account of this argument form requires a general account of when this reasoning is or is not applicable. For real generality, such an account would have to tell some story about what kind of work can be done to pay off the cost of some underdetermination and why we are justified in eliminating the underdetermination otherwise. And these questions are indeed the focus of most accounts of this argument pattern (Fn. 2). Broadly, any legitimate postulate must make some physically significant difference. For Jenann Ismael and Bas van Fraasen this means it must make a difference to quantities "whose values make some discernible impact on gross discrimination of colour, texture, smell, and so on, [no] matter how attenuated the connection is, how esoteric the impact, or how special the conditions under which it can be discerned" (2003, 376). Dasgupta (2016) requires some detectable difference. Perhaps the most liberal analysis is Laura Ruetsche's: "physically significant is significant for physics" (2011, 12).

I don't plan to offer any wholesale account of how to discriminate the superfluous from the physically significant because retail arguments will suffice for my purposes. Indeed, in this section I am only concerned to show that quotienting removes symmetry and therefore underdetermination, and for this we needn't even take a stand on whether that symmetry *ought* to be eliminated. The point is just that it is possible. When it comes to symmetry properties and symmetry relations, this point has already been made by Ian Hacking and Gordon Belot. My contribution in this chapter is to generalize it to count-noun symmetries in a novel way. My contribution in the rest of this dissertation is to show how this generality can be useful.

The point of departure of Hacking's and Belot's discussions is the possible underdetermination associated with the Principle of the Identity of Indiscernibles (PII). This principle asserts that indistinguishable individuals are identical.[14] Thus, it is an anti-underdetermination principle: if it were false then there would be at least two objects that could never be distinguished by any description. A counterexample to the PII would therefore be an example of underdetermination, and so we might try to construct such a counterexample by appealing to symmetry. And indeed, Max Black does just this.[15] It seems

> logically possible that the universe should have contained nothing but two exactly similar spheres.... We might suppose that each was made of chemically pure iron, had a diameter of one mile, that they had the same temperature, colour, and so on, and that nothing else existed. Then every quality and

[14] Of course, "indistinguishable" here requires analysis. See Swinburne (1995) for a menu of options.

[15] For a more general discussion of the debates over this principle and the various moves available on either side, see Hawley (2009).

relational characteristic of the one would also be a property of the other (Black, 1952, 156).

If this world is possible then the PII is false, since there are indistinguishable but non-identical objects.

The indistinguishability of Black's spheres follows from the symmetry of the imagined universe. And such a world does seem conceivable. But, as Hacking (1975) argues, the defender of the PII needn't deny that such a world is conceivable; they need only reject this particular characterization of it. In Black's telling, the spheres must be distinct because "[e]*ach* will have the relational characteristic *being at a distance of two miles*, say, *from the centre of a sphere one mile in diameter*, etc." (1952, 157). And since no object can be two miles from itself, there must be two spheres. Except, as Hacking (1975) observes, an object *can* be two miles from itself, if it lives in a cylindrical universe with a radius of two miles. So whether Black's spheres (or sphere) is a counterexample to the PII depends on which characterization we choose, and we might as well choose the characterization consistent with our position on the PII. "In short, it is vain to contemplate possible spatiotemporal worlds to refute or establish the identity of indiscernibles" (Hacking, 1975, 249).[16] Any description of a symmetric world can be reinterpreted as a redundant description of a non-symmetric world, so any putative counterexample to the identity of indiscernibles can be interpreted away.

Hacking's point generalizes from symmetry properties to symmetry relations, as Belot (2001) observed. This means we can draw a similar moral about putative counterexamples to the Principle of Sufficient Reason (PSR). For example, consider the argument Leibniz levels against the Newtonian conception of absolute space in his correspondence with Clarke.[17]

[16] For more detailed discussion of Hacking's argument, see Landini and Foster (1991), French (1995), and Adams (1979).

[17] Leibniz presents this argument in paragraph five of his third letter the Clarke, for which see Leibniz and Clarke (2000). I will refer to passages in this correspondence by, e.g., "L.III.5" for Leibniz's third letter, fifth paragraph.

Newton's physics describes the behavior over time of objects living in a three-dimensional space. According to this theory as Leibniz and Clarke understand it, there are infinitely many ways for two objects a and b, indistinguishable except by position, to be arrayed in this space at a given time. In particular, for any world containing a and b there is another world that is identical save for the fact that a and b have traded places. Since the only fact about distance in these worlds is the distance between a and b, the two worlds related by a swap are symmetric with respect to the distance facts. But, as Leibniz says, "[s]pace is something absolutely uniform, and without the things placed in it, one point of space absolutely does not differ in any respect whatsoever from another point of space" (L.III.5). On Leibniz's accounting (and Clarke's, C.II.1) distance facts are the only absolute differences between ways the world could be and thus the only reason-giving features in Newton's theory. For Clarke this is evidence of God's free action: the only reason that one world comes about instead of the other originates in God's will.

But for Leibniz this situation is a *reductio* of Newton's theory and reason to revise it. According to Leibniz's PSR, "there is nothing without a sufficient reason why it is, and why it is thus rather than otherwise" (L.III.7). This "nothing" includes God's decisions, so if Newton's account of spacetime were right then it would conflict with the PSR: any world containing at least two objects has a partner that's symmetric with respect to the distance facts, obtained by swapping two objects. God could have no reason to choose one world over a world where two objects are swapped. Thus the PSR in this context is an anti-underdetermination principle, like the PII. Any nontrivial symmetry with respect to the reason-giving features of the world is in conflict with the PSR because it implies the existence of distinct worlds with no differences on which God could make a choice. That is, God's choice would be underdetermined by facts about the worlds. Since Leibniz takes the PSR as axiomatic he must reject Newton's theory for failing to meet this constraint. In its place he argues for a theory on which distance facts are the only physical facts at all. On this theory, once you have specified how far apart a and b are you have specified everything about the world. As such Leibniz's revised theory no longer has any nontrivial symmetries, because no two distinct worlds can be symmetric with respect to the distance facts.

Belot (2001) argues that putative counterexamples against the PSR must be inconclusive by jointly generalizing Hacking's and Leibniz's argumentative moves. Whenever we are presented with symmetric worlds that lead to insufficient reason, we can recharacterize the example as offering two different descriptions of the same world. So Belot's technique, like Hacking's, involves eliminating the underdetermination that threatens the PSR by eliminating the symmetry that leads to it. This works in the swapping case of the previous paragraph, and it works just as well if you'd like to deny the possibility of worlds differing only by a translation in space and time, or by a boost, or some other symmetry.

The further generalization of this technique to count-noun symmetries is rarely much discussed. Or, more accurately, it is taken to coincide with the version applicable to symmetry relations. The recipe for removing count-noun symmetries consists of two steps: identify anything related by the action of the group of symmetries, then define properties, relations, and functions by a rule of supervaluation. I will call this the *strong quotient*, to contrast with the quotienting process I prefer. In full detail, this process is the following.[18]

> **The strong quotient.** Let X be a set and let G be a symmetry group acting on X. We construct the following replacements for X and for the predicates, relations, and functions on X:
>
> 1. Consider the smallest equivalence relation on X such that $x \sim g \cdot x$ for all x in X and g in G. Let $[x] = \{y : y \sim x\}$, and let X/G be the set $\{[x] : x \in X\}$.

[18] Something close to this process is explicitly described by Belot (2001, 2003), Caulton (2015, 156) and Dewar (2017). For the most part these details are left implicit, and it is assumed that quotienting is this process or something close to it. This description of the strong quotient is verbose in the service of clarity. Step 2 is a special case of step 3, since a predicate is just a 1-place relation. And step 3 can be assimilated to step 4 by invoking the product action of G on the n-fold product of X and treating a relation as a map valued in the set $\{\mathsf{true}, \mathsf{false}\}$. The whole construction can be put more concisely by saying that X/G is the coequalizer of the action $G \times X \to X$ and the projection map $G \times X \to X$. This description generalizes immediately to other categories, and this generalization is sometimes the quotient process that gets tacitly appealed to—see, e.g., Butterfield (2006).

2. For any predicate P such that

$$P(x) \text{ if and only if } P(g \cdot x)$$

for all x in X and g in G we have a predicate $P_{/G}$ on X/G defined by

$$P_{/G}([x]) \text{ if and only if } P(x)$$

All other predicates on X are dropped.

3. For any n-place relation R on X such that

$$R(x_1, \ldots, x_n) \text{ if and only if } R(g_1 \cdot x_1, \ldots, g_n \cdot x_n)$$

for all $x_1, \ldots, x_n$ in X and $g_1, \ldots, g_n$ in G we have an n-place relation $R_{/G}$ on X/G defined by

$$R_{/G}([x_1], \ldots, [x_n]) \text{ if and only if } R(x_1, \ldots, x_n)$$

All other relations on X are dropped.

4. For any set Y and any function $f : X \to Y$ such that

$$f(x) = f(g \cdot x)$$

for all x in X and g in G we have a function

$$f_{/G} : X/G \to Y$$

such that $f_{/G}([x]) = f(x)$ for all x in X. All other functions on X are dropped.

Given some count-noun symmetries on a set X, the strong quotient identifies any objects bearing the symmetry relation induced by the class of count-noun symmetries, and it defines relations on the new objects by supervaluation. That is, it eliminates any properties, relations, and structures that are variant under the symmetries, and it turns some invariant properties, relations, and structures on the original representatives into properties, relations, and structures on the new representatives. This means that we can continue to talk about the properties of the original representation that are invariant under the symmetry group, since these correspond to unique properties on the new representatives. This process depends only on the symmetry relation generated by the symmetry group. So, in particular, if two classes of count-noun symmetries on some collection induce the same symmetry relation on the objects of that collection then the results of quotienting by either class coincide. This is the quotienting procedure found throughout the philosophical literature.

As an illustration of the strong quotient, consider the status of absolute velocity in Newtonian dynamics. As Galileo argued, only relative velocities are detectable. You can verify this with some experiments on a ship:

> Shut yourself up with some friend in the main cabin below decks on some large ship, and have with you there some flies, butterflies, and other small flying animals.... With the ship standing still, observe carefully how the little animals fly with equal speed to all sides of the cabin.... When you have observed all these things carefully..., have the ship proceed with any speed you like, so long as the motion is uniform and not fluctuating this way and that. You will discover not the least change in all the effects named, nor could you tell from any of them whether the ship was moving or standing still. ... [T]he butterflies and flies will continue their flights indifferently toward every side, nor will it ever happen that they are concentrated toward the stern, as if tired out from keeping up with the course of the ship, from which they will have been separated during long intervals by keeping themselves in the air.... The cause of all these correspondences of effects is the fact that the ship's motion is common to all the things contained in it, and to the air also (Galilei, 1967, 186–187).

According to the method of symmetry, this experiment—along with the others Galileo describes—gives us reason to think that there is no fact of the matter about absolute velocities. If all the matter in the world were set into uniform motion then it would be just as if it were the interior of a ship's cabin: the results of every experiment would be the same before and after the boost. But if the whole universe is the inside of the ship then there is no outside of the ship, and so it's hard to see what physical facts could differ between the universe before and after the boost.

Supposing that we are moved by the method of symmetry in this case, we should think that two worlds can't differ by a uniform velocity boost. The usual description of Newtonian systems therefore involves some redundancy, since it can describe two systems that differ only in their absolute velocities. The strong quotient is designed to remove just this redundancy and nothing more. More formally, we have some set W of models of worlds that satisfy the laws of Newtonian mechanics, and there is a group B of velocity boosts that acts on W. An element b of B corresponds to a boost with some magnitude and direction, and this element sends a model w in W to the model $b \cdot w$ obtained by uniformly increasing the velocity of all the objects in w by that magnitude and in that direction. Applying the strong quotient gives a set W/B, and an element $[w]$ of this set is an equivalence class of models that differ only by such a boost. The resulting models no longer represent objects as having absolute velocities. For example, these models disagree about the absolute velocity of particle p, so the predicate "has velocity 2 km/s in some direction" will not satisfy the condition in step 2 of the strong quotient and must be dropped. By contrast, the facts that don't depend on absolute values are retained. For example, if some particle p is moving 2 km/s faster than some particle q in some direction according to the model w, then p will also be moving 2 km/s faster than q in the same direction in any equivalent model.

The use of the strong quotient by Hacking and Belot shows that it's always possible to eliminate underdetermination that arises from symmetry. I take it that this is why relatively little attention has been paid to this phase of the method of symmetry. After all, if the strong quotient process always works, then there's really just two kinds of interesting questions about the method of symmetry. There are "pre-quotient" questions about which underdetermination we ought to eliminate, and there are "post-quotient" questions about how we ought to interpret the theory modeled by the quotient structure. There are general pre-quotient questions, like the ones about physical significance from the beginning of this section, and there are specific pre-quotient questions about whether specific instances of underdetermination in specific theories ought to be eliminated (e.g., Healey, 2009, 2.2). And the post-quotient questions motivate a good chunk of the literature in philosophy of physics.[19] These questions aren't completely distinct, but they do presuppose that the quotienting procedure described above is unproblematic.[20] And I disagree.

1.3 Why quotienting doesn't work

I claim that the quotienting procedure of the previous section is not the correct generalization of the quotient procedure to count-noun symmetries. As I stressed in Section 1.1, a class of count-noun symmetries involves more than the symmetry relation it induces: different classes can induce the same relation. Applying the strong quotient eliminates the symmetry, and thus the underdetermination, associated with this relation. And this also eliminates the count-noun symmetries that induce the relation along with the underdetermination that accompanies them. But if we remove just the count-noun symmetries, we should expect to remove *less* than if also remove the symmetry relation they induce. And we're right not to expect this: when we look at the details, the traditional quotienting process returns the wrong verdict in familiar cases. In particular, it removes too much. My alternative quotienting process rectifies this problem.

Let me spell out the heuristic point about what we ought to expect. As we saw in Section 1.1, symmetries lead to underdetermination. And as we saw in Section 1.2, quotienting is a tool for removing symmetries and the underdetermination they generate. So, plausibly, removing more symmetry should remove more underdetermination. But the quotient procedure of the previous section is insensitive to degrees of symmetry and, therefore, to degrees of underdetermination. For example, consider again the snowflake

[19] For examples of post-quotient questions in spacetime theories, see Earman (1989), Saunders (2013), Knox (2014), Weatherall (2015), Wallace (2017), and Dewar (2018). For examples in Yang–Mills theory, see Lyre (1999, 2004), Redhead (2003) and Healey (2007).

[20] In particular, some positions on the pre-quotient questions rely on answers to the post-quotient questions. For example, French's (2014) argues that underdetermination can only be justifiably eliminated if one can answer "Chakravartty's Challenge" by giving "a clear picture" (Chakravartty, 2007, 26) of what the post-quotient models say. Møller-Nielsen (2017) makes similar demands.

of Fig. 1.1. If I ask you to guess how I rotated the snowflake while you were outside, you don't stand a chance. Any visual information you have will be compatible with the rotation I did, but it will also be compatible with my rotation followed by a 60° rotation, or a full rotation, or two full rotations, and so on. So there are infinitely many possibilities compatible with your information. But if I only ask you to guess the net rotation your chances are one in six. So the six-element set of net rotation symmetries leads to less underdetermination than the infinite set of rotation symmetries. However, the quotient procedure of the previous section does not discriminate between these symmetries: they induce the same symmetry relation, so the quotient procedure removes the same amount of underdetermination whether it's the quotient by the infinite set or the six-element one.

But sometimes we want to remove some symmetry but not all. Here is an example. A Euclidean motion is a symmetry made up of some combination of translations and rotations. To specify a translation you need to specify a direction and a distance, and to specify a rotation you need to specify an axis of rotation and an angle of rotation. So to specify a Euclidean motion you need to specify some sequence of direction–distance and axis–angle pairs. Now by Leibniz's argument from the previous section you might be convinced that worlds related by a translation ought really to be counted the same world. But you might like to reserve judgement about worlds related by rotations.[21] So you need a quotienting process that gets rid of the translation symmetry and leaves the rotation symmetry alone.

But that's not the quotienting process on the table. Quotienting out the translation symmetry with the strong quotient leads to absurdity: either a world is rotationally symmetric around infinitely many axes or none. This breakdown is due to the connection between the count-noun sense of symmetry and the property sense that was discussed in

[21] Since this is a hypothetical we can just stipulate that we want to eliminate translation but not rotation symmetries. But this isn't an idle example; this position has recently been defended by Saunders (2013), Knox (2014), Weatherall (2015), Wallace (2017), and Dewar (2018).

Section 1.1. For example, consider any Euclidean motion r consisting of a single rotation, which is therefore characterized by an axis of rotation and an angle of rotation. This motion is a symmetry with respect to all of the observable facts: applying r to any world produces an observationally equivalent world. For some worlds, but not all, this motion is also a symmetry of a stronger sort—namely, it is a symmetry with respect to the facts about which points of space are occupied. For example, consider some world containing a single particle. If this particle is on the axis of rotation of r, then applying r does not change which points of space are occupied. After all, a rotation only moves the points that are not on the axis of rotation. But if the particle is not on the axis of rotation of r, then applying r does change the location of the particle and, therefore, which points of space are occupied. Similarly, applying r to a world containing a ring of matter will leave the occupation facts invariant if and only if the axis of rotation is also the axis of the ring. And, of course, the motion r won't be a symmetry with respect to the occupation facts of any world containing a distribution of matter that doesn't have a rotational symmetry property.

In the pre-quotient theory a world can be rotationally symmetric around some axis without being rotationally symmetric around others. Consider any other rotation r' that has the same angle of rotation as r and a parallel axis of rotation. Before we quotient out the translations there will be very few worlds for which r and r' are both symmetries with respect to the occupation facts. Among the single-particle worlds, the rotation r will be a symmetry with respect to the occupation facts if and only if the particle lies on the axis of rotation of r. It follows that if r is a symmetry with respect to the occupation facts then the particle is not on the axis of rotation of r', since r and r' have parallel axes of rotation. So r' is not a symmetry with respect to the occupation facts of this world. Similarly, the rotation r will be a symmetry with respect to the occupation facts of a world containing a ring of matter around its rotation axis, but r' will not.

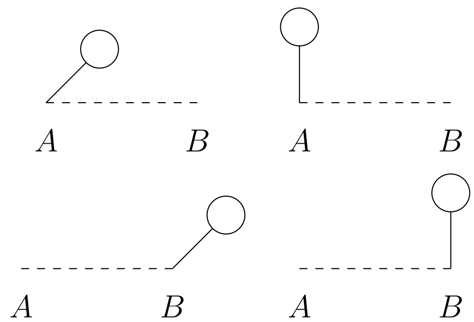

Figure 1.2: Rotation and translation. Two points A and B are 1 m apart. Rotating 45° counterclockwise around A and then translating 1 m to B has the same outcome as translating 1 m to B and then rotating 45° counterclockwise around B.

But after we apply the quotient, any rotational symmetry implies infinitely many others. In particular, the rotation r will be a symmetry with respect to the occupation facts of a world in the quotient structure just in case the rotation r' is as well. To be explicit, let R be the property applying to all worlds for which r is a symmetry with respect to the occupation facts, let R' be the analogous property for r', and let w be some world such that $R(w)$. Since the axes of rotation of r and r' are parallel, they differ by a translation t; let w' be the world obtained by applying t to w. Since r and r' have the same angle of rotation and their axes differ by the translation t, applying r and then translating by t gives the same world as translating by t then applying r'; see Fig. 1.2. By hypothesis $R(w)$, so it follows that $R'(w')$. By definition of $[R]$ we have $[R]([w])$. Similarly, $[R']([w'])$. And since w and w' belong to the same equivalence class this means that $[R']([w])$. Following the linguistic prescription of the quotienting process, we conclude that r and r' are both symmetries of the world $[w]$ with respect to the occupation facts.

Applying the supervaluation rule, we find that any world in the quotient theory has no rotational symmetries with respect to the occupation facts or infinitely many. There are very few worlds for which r and r' are both symmetries with respect to the occupation facts, and when they aren't both symmetries the supervaluation rule says that neither is. But this is surely the wrong conclusion. A world that contains a single particle is rotationally symmetric around that particle and not around any other point of space. So the quotient procedure of the previous section can't be right.

The position I'm advancing in this dissertation is meant to rectify problems like this. The right answer in this case is clear, and the quotient process of the previous section doesn't deliver it. So we should be cautious in contentious cases. Without a general story about how to eliminate underdetermination due to count-noun symmetries, the method of symmetry can't do much for us in the cases we care about. So we need a better quotienting process for the positive project of applying the method of symmetry. A better quotient process would also underwrite or defeat the criticisms of the method of symmetry that I alluded to in the introduction.

1.4 Introducing count-noun identities

As I said in the introduction, we can solve the problem of the previous section by taking numerical identity to have a structure parallel to symmetry. Just as there are symmetry properties, relations, and count nouns, there are identity properties, relations, and count nouns. And just as count-noun symmetries imply symmetry relations imply symmetry properties, count-noun identities imply the identity relation imply self-identity. The amended quotient ought to respect this structure, if it is going to distinguish between classes of count-noun symmetries that induce the same symmetry relation. That is, the correct quotient process should turn count-noun symmetries into count-noun identities, symmetry relations into the identity relation, and symmetry properties into the self-identity property. This gives us a systematic way to avoid problems like the one encountered in the previous section.

Focusing on self-identity and the identity relation might give the impression that our problems are to be found elsewhere. The self-identity property is utterly trivial: every object has it. The numerical identity relation also has the scent of triviality; as Bishop Butler had it, "[e]very thing is what it is, and not another thing" (Butler, 1736).These trivialities might lead you to think that there's nothing more to be said about identity.[22] But every relation is trivial in this way. Every relation holds when it holds and doesn't when it doesn't. Part of the apparent triviality of identity is just this trivial feature of relations. So we shouldn't be too quick to think that anything to do with identity will be trivial.[23] In particular, we have no reason to expect everything to stay so trivial when we go beyond properties and relations.

We saw in the previous section that the strong quotient described in Section 1.2 doesn't work when we are interested in different classes of symmetries that induce the same symmetry relation. In my view, there are problems with both components of this recipe, and so to fix it we have to repair both of them. These problems are rooted in a common presumption that the only relation up to the task of representing identity is identity itself. But this can't be right. For one thing, it's hard to see what could make identity so special in this regard. The "hotter than" relation between objects with temperatures is often represented by the "greater than" relation on real numbers, but it can also be represented by the "less than" relation on real numbers, as on the Delisle or the (original) Celsius scale (Chang, 2004, 160). It can even be represented by a disjunctive combination of the two, as in Tatyana Ehrenfest-Afanassjewa's (1925) treatment of negative temperatures, which uses negative numbers to represent temperatures hotter than those represented by positive numbers. So there are plenty of different relations that can represent the "hotter than" relation; there's no reason to think that the identity relation is any different.

It's also just a matter of fact that we regularly represent identity with other relations. Rational numbers are a simple case of this. A rational number is one that can be represented

[22] Lewis (1986, 192–193) holds this emphatically.

[23] Besides, Martin-Löf (1975) has stuck us with the conceptual possibility of count-noun identities. So really the only question is whether they have any use outside of certain formal systems.

by a ratio of integers p/q with q nonzero. But if some number is represented by p/q then it's also represented by $2p/2q$, and $3p/3q$, and so on. So p/q and r/s represent the same rational number if $ps = qr$, in which case we might say that the pair (p, q) has the same ratio as the pair (r, s). When we use ratios to represent rational numbers we are only concerned with the properties and constructions that do not distinguish between pairs that have the same ratio, since we are only concerned with ratios qua representatives of rational numbers. According to the attitude of Section 1.1 this practice involves some underdetermination, since our manipulations and descriptions don't distinguish between the pair $(1, 2)$, say, and $(2, 4)$. And in high church treatments of the rationals we do apply the strong quotient of Section 1.2 (Enderton, 1977, 102). But this is fastidious; when we actually work with rational numbers we regularly use the "has the same ratio as" relation to represent identity, and it does a perfectly good job.

Both steps of the strong quotient rely on this presumption about how identity must be represented, so both have to be altered. To amend the "identify" step, note that any quotient process that sufficiently discriminates between classes of count-noun symmetries must output more than just a set of possibilities. This follows from the fact that for any structure A there are more classes of count-noun symmetries than there are ways of identifying objects of A. For any way of identifying objects of A there is an equivalence relation such that $a \sim b$ if a and b get identified. And for any equivalence relation there is a class of symmetries inducing that relation: the class of all permutations $\phi : A \to A$ such that $a \sim \phi(a)$. This shows that there are at least as many classes of symmetries as there are ways of identifying objects of A. To see that there are more, note that—as we saw in the case of the snowflake—the observably-equivalent relation is induced by distinct classes of count-noun symmetries. So if the quotient process simply identifies objects of A then it doesn't have enough resources to treat all the classes of count-noun symmetries differently. So to distinguish between different classes of count-noun symmetries, the quotient process must not forget the difference between two symmetries that induce the same functional relation on some part of the collection.

So the quotient process must produce more than equivalence classes. To get a sense of how much more it must produce, consider the simple example of a collection with only one element but more than one symmetry. There is only one functional relation on this collection, so if the quotient is going to distinguish between different classes of symmetries there is no choice but to simply remember the class of symmetries. We should eliminate the symmetries themselves—after all, this is the whole point of quotienting—but the residue of this elimination should allow us to determine whether it is the result of eliminating one symmetry or one hundred. And this amounts to producing a class of count nouns in addition to the one-element collection. Similarly, if a two-element collection has a large class of symmetries, then the quotient process should identify the two elements just in case they are related by a symmetry and give a class of count nouns corresponding to the symmetries that were eliminated.

The upshot of these considerations is that the corrected quotient process has to output a count noun for every count-noun symmetry that it takes as input. This is worrisome. After all, the quotient process is meant to eliminate superfluous structure, and you might think that it can't do this if it also has to remember all of the count-noun symmetries. But this worry turns out to be ill-founded. I show in the next section that there's a precise sense in which the quotient process I am developing here forgets structure, just as it is supposed to. This is possible because it reinterprets the count-noun symmetries. Before the quotient process, count-noun symmetries relate distinct objects that are indistinguishable with respect to some particular properties and relations. The count nouns produced by the quotient process, however, do not relate distinct objects.

I will call the output of the corrected quotient process a "space", in the sense of "logical space".[24] That is, it is "a certain mathematical construct used to represent certain conceptual interconnections" (van Fraassen, 1970, 104). A space in this sense consists of a collection of objects and a collection of identities between the objects of the space. The space of fractions is a paradigm example. An element of this space is a pair (p, q) of integers with q nonzero. For every natural number n and every element (p, q) of the space, there is an identity from (p, q) to (np, nq), also called n. And the quotient of some collection by a class of symmetries gives a space whose elements are the elements of the collection and whose identities correspond to the elements of the class of symmetries.

Spaces allow us to replace the first step of the strong quotient. Rather than producing a set of equivalence classes, the quotient produces a space by adding identities. As I said above, this difference is idle unless we also replace the supervaluation rule with something that does not presuppose the necessity of equivalence classes. The supervaluation rule is a way to turn symmetry-invariant properties, relations, and structures on elements of some collection into properties, relations, and structures on equivalence classes. Since we have replaced the set of equivalence classes with a space, we need to know what a property, relation, or structure on elements of a space looks like, and we need a rule for turning symmetry-invariant properties, relations, and structures on elements of a collection into properties, relations, and structures on elements of the space produced by the quotient.

Properties and relations on elements of a space are the properties and relations that respect the identities. If X is some space and x is an element of X that has property P, then x has property P *qua* element of the space X if any other element y of X that's connected to x by an identity also has property P. In other words, when it comes to properties of elements of a space we apply the identify-and-supervaluate strategy, where two elements are identified just in case they are related by an identity. Indeed, this is the reason that identities in a space X are so-called. The supervaluation rule says that for any

[24] This terminology is also meant to suggest "space" in the sense of Lurie (2009, 1.2.16).

relation R we have $R(x, y)$ if $R(x', y')$ for all x' connected to x by an identity and all y' connected to y by an identity. It follows that the least reflexive relation on elements of X is the relation induced by the count-noun identities of X. Since the identity relation is tautologously the least reflexive relation, two elements x and y of X are identical *qua* elements of X if they are related by a count-noun identity.

As an example of a relation on elements of a space, consider the space of rational numbers. Say that a rational number (p, q) is less than another rational number (r, s) if ps is less than qr as integers. This is a well-defined relation on elements of the space of rational numbers because for all natural n and m we have that (p, q) is less than (r, s) if and only if (np, nq) is less than (mr, ms). And two pairs (p, q) and (r, s) represent the same rational number—they are identical *qua* rational numbers—if $ps = qr$.

The difference between spaces and sets of equivalence classes comes about when we consider maps between spaces and structures on elements of a space. A map from one space to another is a map that sends the elements of the first space to elements of the second *and* identities in the first space to identities in the second. It follows that if the inputs to the map are identical elements of the space then the map gives identical outputs. Take the space of rationals as an example again. The "doubling" map sends a rational (p, q) to the rational $(2p, q)$, and it sends the identity n between (p, q) and (np, nq) to the identity n between $(2p, q)$ and $(2np, nq)$.

We can now state our replacement for the supervaluation rule, giving a complete replacement of the strong quotient. We begin with some collection X that has some class of symmetries. Rather than taking equivalence classes, we form a space by adding identities to the collection X: one identity from x to x' for each symmetry that sends x to x'. The symmetry-invariant properties and relations on elements of the collection X then induce properties and relations on the space. Finally, for any other collection Y with the same class of symmetries and any map $\phi : X \to Y$ that respects the symmetries we have a map between the spaces produced by the quotient that acts as ϕ on the elements of the spaces and as the identity on the identities in the spaces. I make this construction fully precise in the next section.

Before moving on to the formal development of this proposal I show that it resolves the problem with rotation symmetry from Section 1.3. Recall that we were concerned with the collection of all possible distributions of matter in a Newtonian world, and we were interested in the class of symmetries consisting of Euclidean motions—i.e., combinations of translations and rotations. It seems possible for some worlds to have distributions of matter that are rotationally symmetric around some axis. These would be worlds for which there is some set of rotations that produces a distinct world that agrees on the occupation facts. But, as we saw in Section 1.3, the strong quotient forced this set of rotations to be too big.

The quotient process I have described avoids this absurd conclusion. For any world w in our collection there is a set E_w of motions that are symmetries with respect to the occupation facts. We would like to assign a similar set of motions to the space of Newtonian worlds that's obtained by turning translations into identities. To do so, we note that the collection of all translations is also a symmetry of the collection of rotations: for any translation t and rotation r with axis of rotation a and magnitude of rotation θ there is a rotation r' with the same magnitude of rotation but whose axis is obtained by shifting a by t. Therefore, if w is any Newtonian world and w' is the world obtained by shifting w by t, there is a one-to-one correspondence between the elements of E_w and the elements of $E_{w'}$: if r is some rotation that's a symmetry with respect to the occupation facts in w, then shifting its axis of rotation by t will give a symmetry with respect to the occupation facts in w'. So the map sending a world w to the set of symmetries E_w induces to a map of spaces that sends an identity in the space of worlds to the corresponding identity in the space of Euclidean transformations. If some world w is rotationally symmetric about some axis then that rotation will be in E_w, but rotations about other axes may not be. So we aren't led to the problematic conclusion that a world is rotationally symmetric about every axis or none.

This concludes my informal explanation of count-noun identities and their relation to symmetry. To recap: symmetry can be a property, a relation, or a count noun. In each of these manifestations symmetry leads to underdetermination. If we would like to remove this underdetermination, we can do so by removing the symmetry. The usual method for removing symmetry is the strong quotient of Section 1.2. But this recipe is only applicable to certain properties and relations. For example, it misfires when we try to remove translation symmetry from Newtonian worlds. I therefore propose replacing the strong quotient with what I have called the "weak quotient", which involves introducing count-noun identities with a structure parallel to the structure of count-noun symmetries. The weak quotient removes symmetry and underdetermination just as the strong quotient does, but it uses count-noun identities to avoid the problems with the strong quotient. The rest of this dissertation argues that the notion of count-noun identities allows us to make progress on interpretational difficulties in certain physical theories.

1.5 The weak quotient, formally

Actually applying the account of symmetry and identity that I've sketched requires getting more precise on the details, especially if it's to be applied to mathematized theories of physics. In this section I explain one way of making it precise. The formal story I give here makes various simplifying assumptions, but it remains general enough for the interpretational points I want to make it this dissertation. For example, we will suppose that the collection of identities between the elements of every space is just a set, rather than a space.

We will model spaces with groupoids, which consist of a collection of elements and a collection of arrows that satisfy certain properties.[25] We use elements of the groupoid to represent elements of the space and arrows of the groupoid to represent identities between elements of the space. Given a set equipped with a set of symmetries, the weak quotient is be modeled by the action groupoid of the set equipped with symmetries. In the next section I show that the strong quotient is a special case of the weak quotient, applicable for certain properties and relations and for certain sets of symmetries, but not always. Finally, I show that the space produced by the weak quotient generally has less structure than the set we began with and more structure than the set produced by the strong quotient, as I claimed above.

Groupoids are bare-bones mathematical objects with just enough structure to represent count-noun identities.

Definition 1.5.1. A *groupoid* consists of

- a set X of *elements* and
- for all elements x and y of X a set $X(x, y)$ of *arrows* such that
- for all elements x, y, and z and all arrows p in $X(x, y)$ and q in $X(y, z)$ there is an arrow $q \cdot p$ in $X(x, z)$, the *composite* of p and q,
- for all element x there is an arrow id_x in $X(x, x)$, the *identity* arrow of x, such that for all elements x and y in X and arrow p in $X(x, y)$ we have $\mathsf{id}_y \cdot p = p = p \cdot \mathsf{id}_x$, and
- for all elements x and y in X and all arrow p in $X(x, y)$ there is an arrow p^{-1} in $X(y, x)$, the *inverse* of p, such that $p^{-1} \cdot p = \mathsf{id}_x$ and $p \cdot p^{-1} = \mathsf{id}_y$.

[25] More generally we take spaces to be modeled by the $(\infty, 1)$-category of ∞-groupoids. This $(\infty, 1)$-category has various presentations: the Quillen model structure on simplicial sets, the classical model structure on topological spaces, the Thomason model structure on small categories, and more. Since we are restricting attention to spaces in which the collection of identities between any two objects is a mere set, the spaces we speak of in the main text are modeled by 1-coskeletal Kan complexes, homotopy 1-types, and so on.

When there is no chance of confusion we refer to a groupoid by its set of elements. For all groupoid X and elements x and y of X, we write $p : x \to y$ to mean that p is an element of $X(x, y)$.

So a groupoid consists of a set of elements and sets of arrows between those elements such that the "connected by an arrow" relation is an equivalence relation.[26]

Groupoids can be used to model the examples of spaces we saw in the previous section, and they subsume sets. Here are three paradigmatic examples of groupoids to keep in mind.

(i) The space of fractions is a groupoid $\mathbb{Q}$. The set of elements of this groupoid is the set of all pairs of integers (p, q) with q nonzero. The set $\mathbb{Q}((p, q), (r, s))$ is empty unless $ps = qr$. If $ps = qr$ and r/p is an integer, then $\mathbb{Q}((p, q), (r, s))$ is the singleton set containing r/p; otherwise $\mathbb{Q}((p, q), (r, s))$ is the singleton set containing p/r. Composition is multiplication, 1 is the identity arrow, and the inverse of an arrow is its reciprocal.

(ii) For any set X there is a groupoid, also called X, whose set of elements is X and such that $X(x, y)$ is empty unless $x = y$, in which case it is the singleton set containing x. Composition is uniquely defined, and every arrow is its own inverse.

(iii) Recall that a group is a set G equipped with an associative multiplication operation $\cdot$ and an element e such that $g \cdot e = g = e \cdot g$ for all g in G, and such that for every element g there is an element g^{-1} such that $g^{-1} \cdot g = e = g \cdot g^{-1}$. For every group G there is a groupoid BG, the "delooping" of G, whose unique element is the empty set $\emptyset$ and such that $G = BG(\emptyset, \emptyset)$, with composition given by the multiplication operation on G, the element e the identity arrow, and group inverses as groupoid inverses.

[26] In other words, a groupoid is a small category in which every arrow is an isomorphism. For more on groupoids, see Brown (1987) and Riehl (2016, Ch. 1).

Because every set naturally corresponds to a groupoid as in (ii), we will not distinguish between a set and the groupoid it induces.

Action groupoids form another class of examples, and they also implement the first step of the weak quotient described in the previous section. An action of a group G on a set X is a map $G \times X \to X$ that sends a pair (g, x) to an element $g \cdot x$ of X such that $e \cdot x = x$ and $g \cdot (g' \cdot x) = (g \cdot g') \cdot x$ for all elements g and g' of G.

Definition 1.5.2. Let X be a set and G a group of symmetries acting on X. The *action groupoid*[27] $X//G$ of this action is the groupoid with the same elements as X, and such that the set $(X//G)(x, y)$ is the set of elements g of G such that $y = g \cdot x$. Composition, identities, and inverses $X//G$ are given by composition, identities, and inverses in G.

The three examples of the previous paragraph are thus all action groupoids. For any group G its delooping BG is the action groupoid corresponding to the trivial action $G \times \{\emptyset\} \to \{\emptyset\}$. And for any set X the groupoid X is the action groupoid of the action $* \times X \to X$ of the trivial group on X. The positive natural numbers act on pairs (p, q) as $n \cdot (p, q) = (np, nq)$, and $\mathbb{Q}$ is the action groupoid of this action.[28]

To implement the second half of the weak quotient process we must say how symmetry-invariant properties induce properties of the action groupoid. In the case of properties and relations this amounts to supervaluation. Consider some set X with symmetry group G, and let x be an element of $X//G$. For any property P we say that $P(x)$ if and only if $P(y)$ for all y such that $X(x, y)$ has an element. The rule for relations is similar: for any objects x and y of $X//G$ and any relation R we have $R(x, y)$ just in case $R(x', y')$ for all elements x' and y' of X such that $X(x, x')$ and $X(y, y')$ have an element.

As we saw in the previous section, count-noun identities begin to make a difference

[27] Attribute action groupoid to Ehresmann

[28] More precisely, the positive natural numbers form a commutative monoid under multiplication, and this monoid has the action just described. This induces an action of the group completion of the monoid, and $\mathbb{Q}$ is the action groupoid of this group action.

when we consider maps of spaces.

Definition 1.5.3. Let X and Y be groupoids. A map of groupoids consists of

- a set function $f : X \to Y$ on the sets of elements and
- for all elements x and x' of X a set function $f_{x,x'} : X(x, x') \to Y(f(x), f(x'))$ such that
- for any arrows $p : x \to y$ and $q : y \to z$ in X we have $f_{y,z}(q) \cdot f_{x,y}(p) = f_{x,z}(q \cdot p)$,
- for any element x of X we have $f_{x,x}(\mathsf{id}_x) = \mathsf{id}_{f(x)}$, and
- for any arrow $p : x \to y$ we have $f_{y,x}(p^{-1}) = (f_{x,y}(p))^{-1}$.

When there is no chance of confusion we refer to a map of groupoids by its underlying function on elements. We also suppress the subscripts on the functions of arrows. So a map of groupoids $f : X \to Y$ sends an arrow $p : x \to y$ in X to an arrow $f(p) : f(x) \to f(y)$ in Y.

So, for example, a map of groupoids $\mathbb{Q} \to \mathbb{Q}$ is precisely a function of rationals defined in terms of fractions. If X and Y are sets considered as groupoids, a groupoid map $X \to Y$ is just a set function. And for any groups G and G' a map $BG \to BG'$ is a group homomorphism. Finally, for any set X and any group of symmetries G of X there is a map of groupoids $X \to X//G$ that acts as the identity map on objects and for all x in X sends the unique element of $X(x, x)$ to the identity element of $(X//G)(x, x)$.

Until this point we have done little more than restate the strong quotient in terms of groupoids instead of sets. But the point of introducing spaces, and the groupoids that model them, is to go beyond what sets can express. In particular, the weak quotient process can deal with symmetry-invariant maps in a way that sets cannot. Suppose that G is some group of symmetries that acts on two sets X and Y. A set function $f : X \to Y$ is

G-invariant if $f(g \cdot x) = g \cdot f(x)$ for all g in G and all x in X. Any G-invariant set function gives a map $f : X//G \to Y//G$ of groupoids that acts the same way on the elements and acts as the identity on arrows. That is, any arrow $g : x \to g \cdot x$ in $X//G$ is sent to the arrow $g : f(x) \to g \cdot f(x)$ in $Y//G$.

As an illustration, consider again the difficulty of Section 1.3. We had a set W of worlds and a group E of Euclidean motions that are symmetries of W with respect to the occupation facts. We want to say that for every world w of W there is a subgroup E_w of E such that acting on w with an element of E_w produces a world with the same occupation facts. And we would like these symmetries to be preserved if we identify worlds that differ only by a shift. But when we tried this using the strong quotient we were led to absurdity.

If, instead of supervaluating, we turn translation-invariant maps on W into maps of groupoids, then we avoid this problem. Let $\mathsf{Sub}(E)$ be the set of subgroups of E, and let $W \to \mathsf{Sub}(E)$ be the map that sends a world w to the subgroup E_w of symmetries that leave the occupation facts invariant. Since the group of translations is a subgroup $T \subset E$ of the Euclidean motions, we have an action of T on W. We also have an action of T on $\mathsf{Sub}(E)$: for any translation t and subgroup G of E we define $tGt^{-1} = \{t \cdot g \cdot t^{-1} : g \in G\}$. Now if $w' = t \cdot w$ then we have $E_{w'} = tE_w t^{-1}$, since any transformation that leaves the occupation facts of w' invariant will also leave the occupation facts of w invariant when it is shifted by t. Thus the map $W \to \mathsf{Sub}(E)$ is T-invariant, and we obtain a map $W//T \to \mathsf{Sub}(E)//T$. This map sends w to E_w, which is exactly the association we sought.[29]

In sum, the weak quotient modifies both steps of the strong quotient process. The quotient set X/G is replaced by the action groupoid $X//G$, and the supervaluation rule is replaced by a rule that's more appropriate to spaces. In full dress, the process is as follows.

> **The weak quotient.** Let X be a set and let G be a symmetry group acting on X. We construct the following replacements for X and for the predicates, relations, and functions on X:

[29] This is just to say that the stabilizer groups of elements in the same orbit are canonically isomorphic but not equal.

1. Let $X//G$ be the groupoid whose set of elements is X and such that $(X//G)(x, y)$ is the set of elements g of G such that $g \cdot x = y$.
2. For any predicate P such that

$$P(x) \text{ if and only if } P(g \cdot x)$$

for all x in X and g in G we have a predicate $P_{//G}$ on $X//G$ defined by

$$P_{//G}(x) \text{ if and only if } P(x)$$

All other predicates on X are dropped
3. For any n-place relation R such that

$$R(x_1, \dots, x_n) \text{ if and only if } R(g_1 \cdot x_1, \dots, g_n \cdot x_n)$$

for all $x_1, \dots, x_n$ in X and $g_1, \dots, g_n$ in G we have an n-place relation $R_{//G}$ on $X//G$ defined by

$$R_{//G}(x_1, \dots, x_n) \text{ if and only if } R(x_1, \dots, x_n)$$

All other relations on X are dropped.
4. Suppose that Z is a groupoid and $f : X \to Z$ a map of groupoids. Suppose given arrows $p_{g,x} : f(x) \to f(g \cdot x)$ for all x in X and g in G such that $p_{e,x} = \mathsf{id}_x$ and $p_{g,g' \cdot x} \cdot p_{g',x} = p_{g \cdot g',x}$. Then we have a map of groupoids

$$f_{//G} : X//G \to Z$$

such that $f_{//G}(x) = f(x)$ for all x in X and for every arrow $g : x \to g \cdot x$ we have $f_{//G}(g) = p_{g,x}$.

In broad strokes the weak and strong quotients proceed the same way: we replace one collection with another collection with a different identity relation, and then we apply a supervaluation rule. The difference is that the weak quotient replaces the original set with a space, rather than another set, so the supervaluation rule must be applicable to spaces, rather than sets. Since the supervaluation rules of the weak and strong quotients coincide for properties and relations, we should only expect the difference between the weak and strong quotient to matter when we are considering functions on X, not properties and relations.

We have developed the weak quotient in order to make up for shortcomings of the strong quotient. Like the strong quotient, the weak quotient eliminates symmetries—and therefore underdetermination—by producing a new representation in which elements related by a symmetry are identified. The strong quotient also satisfies a stronger requirement, since the representation it produces always represents identity using the identity relation between representatives. So far I've argued that it's possible to relax this stronger requirement, giving the weak quotient. Before moving on to the application of the weak quotient to the interpretation of physical theories, I want to pause to state the relationship between the weak and strong quotients a little more precisely.

1.6 Quotients eliminate structure

It's common to express the "method of symmetry" in terms of structure. Ismael and van Fraassen, for example, try to "identify the sorts of symmetry that signal the presence of excess structure" (2003, 371), which ought to be eliminated. Any quotient process deserving of the name should therefore eliminate structure. In this section I explain a precise sense in which the weak quotient eliminates structure and the strong quotient eliminates more structure than the weak quotient. Appealing to this comparison helps to explain why the strong quotient works as often as it does. The weak quotient reproduces the strong quotient in certain contexts, and it's in these contexts that the strong quotient is successful.

I adopt John Baez and Michael Shulman's (2010, 16) approach to comparing amounts of structure. This approach classifies maps by the amount of structure they forget. We then say that some groupoid E has more structure than a groupoid B if there's some obvious groupoid map $p : E \to B$ that forgets structure. We will show that for any set X

and group G acting on X there are natural groupoid maps

$$X \longrightarrow X//G \longrightarrow X/G$$

which both forget structure, in general. As such, we say that X has more structure than $X//G$ and X/G and that $X//G$ has more structure than X/G. The second map forgets structure unless the action of G on X is free, in which case $X//G$ and X/G coincide. We will also see that the structure forgotten in the move from $X//G$ to X/G makes no difference to the properties or relations of elements of $X//G$; the difference in structure only appears when we consider maps between or structures on groupoids.

The motivating examples for this structure-comparison framework come from algebra. A model of an algebraic theory is a set equipped with some further structure, and the map that forgets the further structure in this case is a paradigmatic groupoid map that forgets structure. So, for example, a group can be modeled by a triple $(A, \times, e)$ of a set A, an associative multiplication operation $\times$ on A, and an identity element e of A. An isomorphism of groups $h : (A, \times, e) \to (B, \star, f)$ is a bijective set function $h : A \to B$ such that $h(a \times a') = h(a) \star h(a')$ and $h(e) = f$. Groups and their isomorphisms assemble into a groupoid Grp: an element of this groupoid is a triple $(A, \times, e)$ and an arrow $(A, \times, e) \to (B, \star, f)$ in Grp is a group isomorphism $(A, \times, e) \to (B, \star, f)$. Similarly, the collection of sets and bijections gives a groupoid Set in which an object is a set X and an arrow $X \to Y$ is a bijection. There is an obvious map $\mathsf{Grp} \to \mathsf{Set}$ that sends a triple $(A, \times, e)$ to A and sends a group isomorphism $(A, \times, e) \to (B, \star, f)$ to the underlying set function $A \to B$. This map intuitively forgets the multiplication operation and identity element, and so elements of Grp have more structure than elements of Set.

We formalize the intuition that $\mathsf{Grp} \to \mathsf{Set}$ is forgetful by appealing to the notion of how many choices are required in reconstructing the forgotten structure. If we start with a group $(A, \times, e)$ and apply the functor $\mathsf{Grp} \to \mathsf{Set}$ we obtain A, and we have to make two choices to reconstruct the group structure: we have to choose a multiplication operation $\times$ and an identity element e. However, these choices are of different characters: while there are generally lots of different multiplication operations on a given set, there is only ever at most one element of a set that can act as an identity element for a given multiplication operation. So choosing a multiplication operation involves more choice than choosing an identity element. More formally, we have the following taxonomy.

Definition 1.6.1. Let $f : E \to B$ be a map of groupoids and b an element of B. The *(homotopy) fiber* of f over b is the groupoid $f^{-1}(b)$ in which

- an element is a pair (e, p) of an element e of E and an arrow $p : f(e) \to b$ in B
- an arrow $q : (e, p) \to (e', p')$ is an arrow $q : e \to e'$ in E such that $p' \cdot f(q) = p$.

Definition 1.6.2. Let $f : E \to B$ be a map of groupoids. We say that f

- *forgets stuff* if there is an element b in B, elements x and y in $f^{-1}(b)$, and two distinct arrows $q, q' : x \to y$ from x to y.
- *forgets at most structure* if for all elements b of B and all elements x and y of $f^{-1}(b)$ there is at most one arrow $x \to y$ in $f^{-1}(b)$.
- *forgets at most properties* if for all elements b of B and all elements x and y of $f^{-1}(b)$ there is a unique arrow $x \to y$ in $f^{-1}(b)$.
- *forgets nothing* if for all b in B the groupoid $f^{-1}(b)$ is nonempty and for all x and y in $f^{-1}(b)$ there is a unique arrow $x \to y$ in $f^{-1}(b)$.

We also use various cognates of these: a map *forgets something* if it's not the case that it forgets nothing, it *forgets properties* if it forgets something and forgets at most properties, and it *forgets structure* if it forgets at most structure and it's not the case that it forgets at most properties. Any groupoid map *forgets at most stuff.*

According to this taxonomy both the weak quotient and the strong quotient generally forget structure, and the strong quotient forgets more than the weak quotient does. Let X be a set and G a group acting on X. There are natural maps

$$X \longrightarrow X//G \longrightarrow X/G$$

The first map acts as the identity on objects and arrows: it sends every element x of X to the element x of $X//G$, and it sends every arrow $\mathsf{id}_x : x \to x$ in X (since were considering X as a groupoid) to the arrow $\mathsf{id}_x : x \to x$ in $X//G$. In general, this map forgets structure.[30] Indeed, the only case in which it doesn't forget structure is when G is the trivial group, in which case we have $X = X//G$ and the map $X \to X//G$ is the identity map, which forgets nothing. The composite map $X \to X/G$ also generally forgets structure.[31] So the weak and strong quotients are both ways of eliminating structure. And the map $X//G \to X/G$ generally forgets stuff, so the weak quotient eliminates less structure than the strong quotient does.[32] In other words, the strong quotient is a two-step process: the first step eliminates some structure to produce $X//G$, and the second step eliminates further stuff to produce X/G.

Characterizing the strong quotient as a two-step process invites a question: what does the second step do? When the strong quotient was introduced in Section 1.2 we

[30] *Proof.* For any element x of $X//G$ the fiber of $X \to X//G$ over x consists of the set of pairs (y, g) such that $g \cdot y = x$. There is at most one arrow $(y, g) \to (y', g')$ in this fiber, since there is at most one arrow $y \to y'$ in X, so the map $X \to X//G$ always forgets at most structure. And in general it does forget structure: let G be a nontrivial group, and consider the map $* \to *//G$ given by the trivial action of G on the singleton set. Any distinct elements g and g' of G give distinct elements of the fiber of $* \to *//G$ over the unique element of $*//G$, so this map does not forget at most properties.

[31] *Proof.* For any element $[x]$ of X/G the fiber of $X \to X/G$ over $[x]$ consists of the set of elements y in X such that $[x] = [y]$. Since this is a set there is at most one arrow between any two elements, so $X \to X/G$ forgets at most structure. And in general it does forget structure: let G act nontrivially on some set X, so that there are some elements x and y of X and g of G with $g \cdot y = x$. Then there is no arrow $x \to y$ in the fiber of $X \to X/G$ over $[x]$.

[32] *Proof.* Consider any nontrivial group G acting on the singleton set $*$. Then the fiber of $*//G \to */G$ over the unique element of $*/G$ is the groupoid BG, which has a single element and distinct arrows g and g', since G is nontrivial.

conceived of the move from X to X/G as involving a single step: identifying the elements of X that are related by a symmetry in G. But now that the weak quotient's in view we can see that it's really the move from X to $X//G$ that does this identification; the move from $X//G$ to X/G does something else.

The move from $X//G$ to X/G is also a process of identification. It is an instance of a more general construction, applicable to any groupoid, that identifies any two identities that stand between the same elements of the groupoid.

Definition 1.6.3. The *truncation* of a groupoid X is the groupoid $\|X\|$ with the same elements as X but such that $\|X\|(x, y)$ is the set of equivalence classes of the universal relation on $X(x, y)$—i.e., the equivalence relation such that $p \sim q$ for all $p, q : x \to y$. The truncation of a groupoid map $f : X \to Y$ is the unique map $\|f\| : \|X\| \to \|Y\|$ that acts as f on the elements of $\|X\|$.

For any groupoid X there is a canonical groupoid map $X \to \|X\|$ that acts as the identity on elements and sends every arrow to its equivalence class. In general, this map forgets stuff.[33] And when we quotient a set by a group of symmetries this map fits into the quotient process, giving

$$X \longrightarrow X//G \longrightarrow \|X//G\| \longrightarrow X/G$$

The first map generally forgets structure and the second map generally forgets stuff; the third map always forgets nothing. So $\|X//G\|$ represents the same space that X/G does, and this space is obtained by truncating the weak quotient $X//G$.

The truncation helps to characterize what's eliminated in the move from the weak quotient to the strong quotient. It also helps to explain the relative success of the strong quotient in interpretive contexts. When we restrict attention to properties and relations on a groupoid, the stuff that's forgotten by truncation is irrelevant. A well-defined predicate P

[33] *Proof.* Let G be a nontrivial group and let X be BG. Then $BG \to \|BG\|$ is the map of Fn. 32.

on a groupoid X is a predicate such that $P(x)$ if and only if $P(y)$ for all arrows $p : x \to y$. It follows that any well-defined predicate on a groupoid X is also a well-defined predicate on the truncation $\|X\|$, and vice versa. Two objects in X are connected by some arrow just in case the corresponding objects in $\|X\|$ are connected by some arrow, and since properties don't depend in any way on specific arrows, this means that properties on X are equivalent to properties on $\|X\|$. The same holds for relations on X, for the same reasons.

Thus, as long as we only attend to properties and relations, a groupoid is interchangeable with its truncation. Since the strong quotient is equivalent to the truncation of the weak quotient, the strong and weak quotients will be interchangeable as long as we're only concerned with properties and relations on the quotient. It's only in step 4 of the strong quotient process, when we consider maps on the strong quotient, that the difference between the strong and weak quotient makes a difference. And even then, the difference only matters when the map takes values in a groupoid with non-identity arrows.

The formalization of the last two sections shows that the weak quotient is the appropriate mechanism for removing symmetry and the underdetermination it produces. The different kinds of symmetry in Section 1.1 all lead to underdetermination because they all involve distinctions that we can't track with our referential, empirical, or other capacities. To eliminate the underdetermination it thus suffices to eliminate the distinctions, and this is what the weak quotient does. As we have seen in this section, the strong quotient goes further, eliminating any distinctions between identities. So the method of symmetry only really licenses applications of the weak quotient; the further truncation step lacks justification. Whether this is a real problem depends on what we lose by truncating. Propositions and relations don't depend on identity structure, so truncation makes no difference unless we have some reason to care about other maps or structures on the quotient set. The rest of this dissertation deals with examples of such maps and structures.

1.7 Using count-noun identities

It's difficult to say anything informative and general about when a physical theory's models will involve count-noun identities. It's like trying to say when a physical theory will invoke differential calculus: any informative description of the kinds of phenomena well-described by differential calculus will essentially have to characterize them as "the phenomena well-described by differential calculus". But there are some patterns. Most of the count-noun identities in this dissertation have their origin in differential calculus. It's often the case that we would like to posit some unobservable feature characterized only by its derivative. If we don't allow for count-noun identities in these cases then we cannot introduce these features without also introducing some symmetry and therefore underdetermination. The strong quotient removes this symmetry, but it also removes the postulated feature. But if we allow for count-noun identities then we can have these features without any symmetry. Thus these structures are examples of what truncation eliminates.

Most students of physics first encounter count-noun identities in the theory of energy, where they also learn some conventions to avoid thinking of them as identities. In Newtonian mechanics and gravitation it's often possible to think of a system as possessing an energy, a quantity of motion that is neither created or destroyed. For example, a pendulum demonstrates a cyclic conversion between kinetic and gravitational potential energy. The kinetic energy of an object increases with its speed, and the gravitational potential energy of an object near the earth increases with its distance from the earth. So when the pendulum moves from the top of its swing to its lowest point it gains kinetic energy and loses gravitational potential energy. As it swings back to the top it gains gravitational potential energy while losing kinetic. Similarly, a block attached to a spring gains elastic potential energy when the spring is squashed or stretched, and it oscillates back and forth as this potential energy is converted to kinetic energy and back. In both

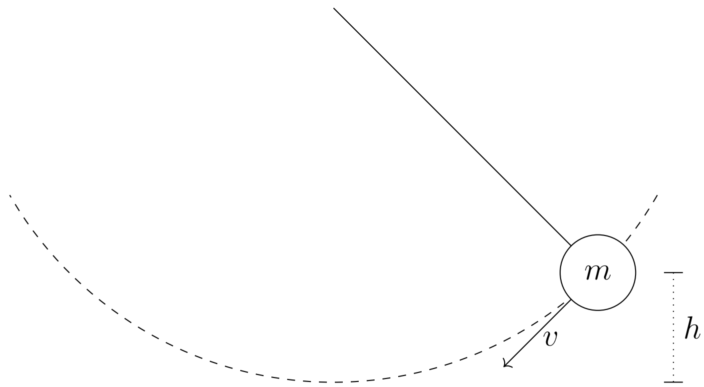

Figure 1.3: A pendulum

the pendulum and the spring, the sum total of kinetic and gravitational potential energy at any moment is the same as the total at any other.

Systems that display this kind of energy conservation are called, naturally, conservative. More precisely, a force F is called conservative if there is some potential ϕ such that $F = -d\phi$. A conservative system, like the pendulum or spring, is a system in which all the relevant forces are conservative. In any introductory physics textbook you'll find a treatment of the pendulum's energetics that goes something like this.[34] Consider the pendulum in Fig. 1.3, whose bob has mass m. The kinetic energy of an object with mass m and velocity v is $\frac{1}{2}mv^2$. The force of gravity on the bob is $F = mg$ downward, with g a constant. So if we take $\phi = mgh$, with h the height of the bob above its lowest point, then we have $F = -d\phi$. The total energy of the pendulum is thus $E = \frac{1}{2}mv^2 + mgh$, and it follows from Newton's laws that this quantity does not change in time. We can give a similar analysis for a block on a spring by taking the elastic potential energy to be $\frac{1}{2}kq^2$, where k is a constant characterizing the spring and q is the displacement of the system from equilibrium. Analogous formulas abound for calculating potential energy in other forms.

[34] The following discussion of a pendulum and spring can be found in any introductory textbook on mechanics. Spivak (2010) is the example closest to hand; he discusses pendulums and springs on pages 93 and 292, respectively.

On one straightforward reading of this analysis the total energy is a quantity associated with the system, and as the pendulum evolves it might convert portions of the energy from kinetic to potential and back. A pendulum at rest at the bottom of its swing has zero velocity or gravitational potential energy, so it has zero energy. If the pendulum is swinging then it will have some velocity v_{max} at the bottom of its swing and some height h_{max} at the highest point of its swing. Since the pendulum conserves energy we must then have $E = \frac{1}{2}mv_{\text{max}}^2 = mgh_{\text{max}}$, so the pendulum's energy is all kinetic at the bottom of its swing and all potential at the top. In between it's some mix of the two.

But taking energy to be a quantity assigned to systems has difficulties. Despite the formulas just given, it's never actually possible to measure or uniquely calculate anything but energy differences. The analysis of the pendulum would go through just the same if the gravitational energy were $\phi' = mgh + 2$, for example, making the total energy of the pendulum $E' = \frac{1}{2}mv^2 + mgh + 2$. It would still follow from Newton's laws that E' does not change over time, that kinetic energy would be maximized and gravitational potential energy minimized at the bottom of the pendulum's swing, and that kinetic energy would be minimized and gravitational potential energy maximized at the top of the swing. Indeed, the pendulum would be swinging at the same speed and to the same height whether the gravitational potential energy were mgh or $mgh + 2$. So if there were some physical difference between ϕ and ϕ' it would be undetectable. "That we can know only energy *differences*—that energy's absolute value would be a 'dangler', making no difference to anything else—suggests (though falls well short of showing conclusively) that there is *no fact* of the matter regarding the absolute value of a system's energy" (Lange, 2002, 128).[35] Of course, this reasoning is just the method of symmetry. It's also a common line; similar claims are found throughout the physical and philosophical literature.[36]

[35] See Lange (2002, Ch. 5) for a fuller discussion of some of the ontological issues surrounding energy, particularly in the context of electromagnetism and the energy associated with fields. The ontological questions of interest to Lange are not exactly the ones we are investigating, but they are close cousins.

[36] Physics textbooks are sometimes evasive on this point. After defining energy, they'll point out that it

It's not immediately clear what it would mean for there to be a fact of the matter about energy differences but not about the absolute value of different systems' energies. We can't interpret this claim too literally, because if systems don't have energies then there can't be differences between systems' energies. It's true that energy differences survive the strong quotient: if we shift the energy of every system by some constant amount then the energy differences between systems are preserved. But the linguistic prescription of the strong quotient misfires. It would be misleading to persist in talking about energy differences when we move to the strong quotient, where there are no longer energies to have differences.

The linguistic prescription of the method of symmetry fails, so if we'd like to eliminate the shift symmetry we need to find some other characterization of the quantitative relation that's represented by energy differences in the pre-quotient models. The biggest problem we face in this task is the fact that energetics are so often applied to isolated systems. For example, the analysis of the pendulum that we just sketched doesn't compare it to some other system. And it's not obvious how any of the formulas for gravitational potential energy, which seemingly depend only on features of the pendulum, can be understood as describing a relation. It would be one thing if these calculations only had significance when there were many systems that could be related to one another; in this case we might be able to understand the analysis just given as a degenerate case of a computation that has physical significance when there are multiple systems. But energy conservation places physical constraints on the pendulum—for example, we must have $\frac{1}{2}mv_{\text{max}}^2 = mgh_{\text{max}}$—so the energetics of the pendulum aren't degenerate.

is only defined up to a constant and that we usually use this ambiguity in the definition to conveniently redefine what counts as "zero energy" from problem to problem (e.g., Spivak, 2010, 89). I don't really understand a sense of definition on which it makes sense for something to be defined "up to a constant". One possibility is that we're meant to combine the given description of energy's theoretical role with an act of ostensive baptism to give an actual definition in any given situation (cf. Kripke, 1980, 54). On this reading it would be possible for two worlds to differ by a global shift in energy, though they would be observationally indistinguishable. The other possibility is something like the one I engage with in the main text: the absolute value of the energy isn't physically significant, but energy differences are.

Count-noun identities give us a way out of this interpretational muddle that generalizes to more complicated cases.[37] In the case of pendulums near the surface of the earth the solution goes like this. The space of possible energies is a one-element space with one count-noun identity for each real number; it is represented by the groupoid $B\mathbb{R}$. Every configuration of the pendulum has some energy, so each configuration is assigned some element of this space of energies. Of course, since the space of possible energies has only one element every configuration is assigned the same energy. That is, the energy assigned to any one configuration is identical to the energy assigned to any other configuration. But there is a further fact of the matter about which identity is instantiated When we are working with groupoids like $B\mathbb{R}$ there's a trick we can use to altogether between the two configurations, and here there are many possibilities: one for every real number. So on this interpretation, energy differences in the pre-quotient models represent count-noun identities.

We've been using groupoids to model count-noun identities, but treatments of potential energy almost never use groupoids explicitly. They're able to avoid groupoids because when we're working with groupoids of the form BG for some group G, we can always "undo" the weak quotient to give a set of possibilities equipped with a set of symmetries. That is, we can pick some set X and some action of G on X such that for all x and y in X there is a unique g in G with $g \cdot x = y$.[38] We can then think of X as the space of possible values, and for any x and y in X we can think of the unique g such that $g \cdot x = y$ as the difference or ratio of x and y. So rather than thinking of every system as being assigned the unique element of BG and any pair of systems as being assigned some identity in BG, we can instead think of every system as being assigned a value in X; the ratio or difference of two values then comes from the action of G on X. We must remember,

[37] The following discussion is an adaptation of Baez's (2009) discussion in terms of torsors into groupoid language.

[38] Such an action is said to be free and transitive. Any group acts freely and transitively on itself by left multiplication, so it's always possible to choose some such action.

however, that the choice of X was arbitrary and that there is no fact of the matter about which element of X is assigned to which system. We can use this arbitrariness for our convenience: picking an element of X to label as the identity fixes a unique isomorphism between X and G, so we can give X a group structure and speak of differences and ratios literally.[39] This is what we do when we pick a conventional "zero energy", identifying the space of possible energies with the space of real numbers. But in introducing a second kind of arbitrariness we also have to keep track of which features are merely conventional and which are physically significant. The only physically significant fact is the group element that connects the two elements of X. This technique allows us to avoid directly modeling count-noun identities, but we have to pay the price of carrying around certain tacit—and possibly subtle—restrictions on which mathematical features we take to have physical significance.

This groupoid-avoidance technique does not directly generalize to all cases of count-noun identities in physics. The example of the gravitational potential energy of a pendulum has a number of special features that allow for the conventions that support the technique: the group $\mathbb{R}$ is topologically contractible, there is a global space–time split, we are considering a simply-connected spacetime, and the gravitational force field is constant throughout space and time. The technique of the previous paragraph turns a space BG with one element and many identities into a space X with many elements and no nontrivial identities (i.e., a set). The fact that the elements of X represent identities in a space with only one element is encoded in the action of G on X, and this encoding relies on the special features just mentioned. In other words, these features allow us to turns identities into elements, remembering only that there is a unique element in the space of interest. So introducing these conventions won't generalize to spaces that have many

[39] This isomorphism comes about as follows. Choose any element x_0 of X. The action then induces a map $G \to X$ that sends g to $g \cdot x_0$. Since the action is free this map is an injection, and since the action is transitive this maps is a surjection. By the fact that the map comes from an action we have $e \cdot x_0 = x_0$ for e the identity of G, and this means that x_0 is the identity element of the induced group structure on X.

elements and many identities, because when we do so we forget how many elements we started with. Generalizations of these conventions do exist—the theory of principal bundles is an example—but we can see the physical meaning of count-noun identities most clearly if we avoid using these techniques.

Potential energy is a simple example of one common source of count-noun identities in physics: antiderivatives. Potential energy is completely characterized by its derivative, and so the method of symmetry suggests that only energy differences have physical significance. Read literally this is absurd, so we need an alternative interpretation of what energy differences represent. Count-noun identities allow for such an interpretation, and what's more they give an explanation of the usual mathematical treatment of energetics. Seeing these features of gravitational potential energy as an expression of count-noun identities is a useful guide to the interpretation of similar phenomena in other theories, with which we now close.

1.8 Yang–Mills theory

The next two chapters focus on count-noun identities in Yang–Mills theory, a generalization of electromagnetism that describes the three fundamental forces in quantum field theory. These chapters focus on specific issues in the interpretation of these theories, and so they don't explain the general picture of Yang–Mills theory I'm offering. That's the goal of this section, which is expository and will presume some familiarity with Yang–Mills theories. The first half of the section explains my account of so-called "gauge symmetries". On my account, these are not symmetries in the sense that I have been using the term; rather, they are count-noun identities. It follows that only gauge-invariant quantities and constructions have physical significance on my account, since they respect the identity relation. While this conclusion isn't uncommon, its consequences haven't been fully recognized. I investigate these consequences further in Chapter 2. I also explain how count-noun identities interact with locality in field theories, which is discussed in further detail in Chapter 3. Since locality conceptually depends on identity, taking gauge transformations to be count-noun identities means incorporating gauge-invariance into the characterization of locality.

To see how count-noun identities arise in field theories we contrast a handful of field theories of electromagnetism in special relativity. Fix Minkowski space as the background spacetime. One theory represents the electromagnetic state of affairs using the electromagnetic field strength F, a closed 2-form.[40] To say that F is closed means that $dF = 0$, where d is the exterior derivative. Since it's a 2-form it sends any vectors u and v to a real number $F(u, v)$, and two 2-forms are equal just in case they are represented by the same array of numbers in some (and therefore any) coordinate system. That is, there is at most one identity between any two configurations. This makes the configuration space of this theory a set, which we call Ω^2_{cl}.[41]

An alternative theory represents the electromagnetic state of affairs with the electromagnetic potential A, a 1-form. Such potentials are usually introduced in a manner analogous to the gravitational potential of the previous section: given some field strength F, the potential is a field A such that $F = dA$. Taking A to be a 1-form means that it has identity conditions similar to F: it sends any vector v to a real number $A(v)$, and two 1-forms agree just in case they are represented by the same array in some (hence any) coordinate system. So there is at most one identity between any two configurations, and the configuration space is again a set, which we'll call Ω^1.[42]

---

[40] I assume familiarity with the theory of differential forms, for which see Baez and Muniain (1994), Hubbard and Hubbard (2001), or Malament (2012). My notation follows Baez and Muniain's, except where otherwise noted.

[41] Something close to this theory is discussed by Baez and Muniain (1994), Jackson (1999, 11.6), Malament (2012, 2.6), and Weatherall (2016c). This differs from the theory Weatherall calls "EM_1" in that I have fixed the background spacetime. It differs from the theory Malament presents in that we assume F is closed, rather than imposing $dF = 0$ as a dynamical equation.

[42] This theory is discussed by Belot (1998, 3.1), who calls it the "vector potential as a physical field"

If we take the electromagnetic configuration space to be Ω^1 we run into problems analogous to the ones we encounter when we take energy to be represented by a real number. The dynamical equations of electromagnetism involve the field strength, not the potential. In the presence of some current j the field strength F must satisfy the equation $\star d \star F = j$, and the force on some test particle of charge q and four-velocity u is $q\,F(u, -)$. In the second theory the equations are the same, taking F to be given by dA. This means that two potentials with the same field strength are dynamically indistinguishable: one satisfies the dynamical equations if and only if the other does, and they have the same effects on all charged matter—our measuring instruments included. So if we take the electromagnetic configuration space to be Ω^1 then we have introduced empirical underdetermination, just as we did by introducing gravitational potential energy.

The dynamical situation is much worse here than in the case of potential energy, however. For any smooth scalar field f we have $d(A + df) = dA$, since $d^2 = 0$. So A and $A + df$ have the same field strength, making them dynamically indistinguishable. This makes prediction impossible: for any possible electromagnetic history A and any time t there are infinitely many alternative histories $A + df$, where f is some smooth scalar field that's constant up until t. These are all dynamically possible, and they are indistinguishable before time t, so knowing the history of the world up to some time does not suffice to uniquely specify the future.

This situation should be familiar by now. If we take the configuration space of electromagnetism to be Ω^1 then we have a large class of symmetries from which follows a large amount of underdetermination. The method of symmetry then advises that we eliminate this underdetermination by eliminating the symmetries. And indeed, it's a matter of textbook electromagnetism that "[o]nly those quantities have physical meaning which are invariant with respect to the transformation" sending A to $A + df$ (Landau and

interpretation. It is similar to Weatherall's (2016c) "EM_2", though I have fixed the background manifold.

Lifshitz, 1980, 53). If we apply the strong quotient then we obtain Ω^2_{cl}, and we're back to the first theory we considered. This suggests that the introduction of the potential is a matter of computational convenience. The symmetry-invariant content of the second theory, according to the traditional analysis, is just exactly the content of the first theory. However, as we saw in Section 1.6, the strong quotient eliminates more structure than is licensed by the method of symmetry. To eliminate the empirical and dynamical underdetermination that accompanies the configuration space Ω^1 we need only apply the weak quotient.

I've been using electromagnetism as a stand-in for Yang–Mills theories more generally. A Yang–Mills theory on Minkowski space is specified by a Lie group G; electromagnetism is the case of $G = U(1)$. The possible configuration spaces of electromagnetism generalize.[43] The electromagnetic potential generalizes to a $\mathfrak{g}$-valued 1-form A, so one option for the configuration space of this theory is the set $\Omega^1(\mathbb{M}^4; \mathfrak{g})$ of $\mathfrak{g}$-valued 1-forms. However, this set is too big, just as the set Ω^1 was too big in the case of electromagnetism. That is, there are distinct elements of $\Omega^1(\mathbb{M}^4; \mathfrak{g})$ that correspond to the same physical state of affairs. In particular, two $\mathfrak{g}$-valued 1-forms A and A' represent the same state of affairs if there is some smooth map $g : \mathbb{M}^4 \to G$ such that

$$A' = \mathrm{Ad}_g(A) - g^*\overline{\theta}$$

where Ad_f is the adjoint representation of G on $\mathfrak{g}$ and $\overline{\theta}$ is the right-invariant Maurer–Cartan form on G.[44] Such a map is called a gauge transformation. In the case of electromagnetism, where G is $U(1)$, the adjoint representation is trivial and the term $g^*\overline{\theta}$ is equal to $-df$ for any f such that $g = \exp(if)$, using the canonical isomorphism of $\mathfrak{u}(1)$ and $\mathbb{R}$. So

[43] Strictly speaking, for a Yang–Mills theory we should also assume that G is either compact and simple or $U(1)$. This ensures that the kinetic term in the Lagrangian is positive-definite (Weinberg, 1996, 15.2). The general picture generalizes to other dynamics, however, and this ignoring this restriction makes no important difference to the kinematics of the theory.

[44] My notation here diverges from Baez and Muniain's (1994), since they don't give a coordinate-free expression for a gauge transformation. I adopt the notation of Schreiber and Waldorf (2009).

the map sending an electromagnetic potential A to $A + df$ for some smooth f is a gauge transformation.

I take the correct configuration space of Yang–Mills theory on the whole of Minkowski space to be the weak quotient of $\mathfrak{g}$-valued 1-forms by the group of gauge transformations. The method of symmetry advises identifying two potentials that are related by a gauge transformation. The weak and strong quotients give three spaces

$$\Omega^1(\mathbb{M}^4;\mathfrak{g}) \longrightarrow \Omega^1(\mathbb{M}^4;\mathfrak{g})//G \longrightarrow \|\Omega^1(\mathbb{M}^4;\mathfrak{g})//G\|$$

The first of these is the set of $\mathfrak{g}$-valued 1-forms on Minkowski space, and the third is the set of gauge-equivalence classes thereof. Philosophical debates over the interpretation of gauge theories assume that these are the only two options, and that the correct interpretation of Yang–Mills theory must produce a configuration space isomorphic to one of these. But, as I argue in the rest of this dissertation, the structure forgotten in the truncation of the weak quotient is physically significant structure. Gauge-equivalent configurations do represent the same physical state of affairs, and so we ought to quotient, but we have no reason to go the extra step and throw out the count-noun identities this quotient produces. In fact, we have positive reason to retain these identities, since they have physical significance.

In the context of Yang–Mills theory the significance of count-noun identities is easiest to see in terms of the process of gauge fixing. The existence of many potentials in $\Omega^1(\mathbb{M}^4;\mathfrak{g})//G$ that represent the same physical state of affairs "gives us the possibility of choosing them so that they fulfil one auxiliary condition chosen by us" (Landau and Lifshitz, 1980, 53). In the case of electromagnetism, for example, it might be convenient to ask for the potential to be in Lorenz gauge, satisfying $\star d \star A = 0$ (in coordinates, $\partial_\mu A^\mu = 0$). For any potential A we can find some f such that $\star d \star A = \star d \star df$, meaning that $\star d \star (A + df) = 0$. Since A and $A + df$ are gauge equivalent they represent the same state of affairs, showing that we can always choose a potential in Lorenz gauge to represent the electromagnetic configuration. If we have chosen some set of coordinates on spacetime other choices might be convenient: we can make A_0 vanish by choosing some f with $\partial_0 f = A_0$, or we can ask for the divergence of the spatial part of A to vanish, among other common possible restrictions. Imposing some requirement like this is called fixing a gauge.

The point of gauge-fixing is that it doesn't forget anything physical. The potentials A and $A + df$ both represent the same electromagnetic state of affairs, so we can always choose to use the latter when it's convenient. More formally, fixing a gauge means replacing $\Omega^1(\mathbb{M}^4; \mathfrak{g})//G$ with a subgroupoid without forgetting anything; potentials that belong to this subgroup are said to be in the chosen gauge. Recall that a groupoid map $f : E \to B$ forgets nothing if for any element b of B the groupoid $f^{-1}(b)$ is nonempty and for all x in y in $f^{-1}(b)$ there is a unique arrow $x \to y$. Equivalently, a map $f : E \to B$ forgets nothing if there is another map $g : B \to E$ and arrows $x \to g(f(x))$ and $f(g(b)) \to b$ for all x in E and b in B. Since arrows represent count-noun identities, this is the well-defined notion of a groupoid map having an inverse. A groupoid map that forgets nothing is called an equivalence, and two groupoids are equivalent if there is an equivalence between them.

On my view, fixing a gauge is an instance of choosing a subgroupoid that's equivalent to its supergroupoid. A groupoid X is a (full)[45] subgroupoid of a groupoid Y if every object of X is an object of Y and $X(a, b) = Y(a, b)$ for all objects a and b in X. So, for example, the groupoid $\mathbb{Q}$ of rationals has a subgroupoid $\overline{\mathbb{Q}}$ consisting of the pairs (p, q) whose greatest common divisor is 1—i.e., such that p/q is written in lowest terms. If X is a subgroupoid of Y then there's a groupoid inclusion map $X \hookrightarrow Y$ that sends each object and each arrow of X to itself. And this map can be an equivalence: the map $\overline{\mathbb{Q}} \hookrightarrow \mathbb{Q}$ is an example of an inclusion that's an equivalence, because there is a unique arrow in $\mathbb{Q}$ between any pair (p, q) and its lowest form. So fixing a gauge is analogous to writing a

[45]explain

fraction in its lowest form. It's a choice of representation that involves no assumptions about the thing being represented.

This understanding of gauge fixing is slightly different from the understanding one sometimes encounters. Fixing a gauge is sometimes treated as though it meant picking a representative from each equivalence class in the truncated quotient. The most obvious way that this picture of gauge fixing differs from mine is its use of the truncated quotient instead of the weak quotient. Gauge fixing can eliminate some representatives, but it can never eliminate count-noun identities. Suppose that X is a gauge-fixed subgroupoid of $\Omega^1(\mathbb{M}^4; \mathfrak{g})//G$ such that for every potential A in $\Omega^1(\mathbb{M}^4; \mathfrak{g})//G$ there is precisely one A' in X such that there's an arrow $A \to A'$. We would then have imposed enough gauge-fixing conditions that there is a one-to-one correspondence between states of the fields and elements of X. But if this is a gauge fixing then the inclusion $X \hookrightarrow \Omega^1(\mathbb{M}^4; \mathfrak{g})//G$ is an equivalence, and so $X(A, A')$ is the set of all gauge transformations from A to A'. So moving to X does not eliminate any of the count-noun identities between configurations, and gauge-fixing can't turn a groupoid into a set if it wasn't one to begin with.

What's more, my account of gauge fixing doesn't require the gauge-fixed subgroupoid to have only one object per physical state of affairs. This is because most conditions that physicists call gauge-fixing conditions don't do this. Careful treatments of gauge fixing are almost always clear about this. For example, consider electromagnetism again and the requirement that the potential be divergence free. As we saw above, for any potential A there is some f such that $A + df$ is in Lorenz gauge. But then so is $A + df + df'$, for any f' such that $\star d \star df' = 0$. And there are infinitely many of these.[46] In principle you could cook up a gauge-fixed subgroupoid by picking one potential from each gauge-equivalence class and forming the subgroupoid of these, but the gauge-fixing conditions encountered in

[46] *Proof.* Choosing some coordinates, this condition on f' becomes $\partial_\mu \partial^\mu f' = 0$. So f' can be any harmonic function, such as $e^x \sin y$.

nature are never of this kind.[47]

The point of all this is that gauge-fixing techniques and conditions have no physical significance, and on my view this follows from the fact that the output of a gauge-fixing procedure is equivalent to the input. The other side of this coin is that any physically significant restriction or construction must not distinguish between equivalent groupoids. This is related to the fact that physically significant properties, relations, and functions must not distinguish between objects of some groupoid that are related by an arrow. For example, suppose that we would like to restrict attention to those configurations that vanish inside some region Σ. Naively, we might form the space of such configurations by starting with $\Omega^1(\mathbb{M}^4; \mathfrak{u}(1))//U(1)$ and throwing out any potentials A whose restriction to Σ is nonzero. But this is physically meaningless; if f is some smooth function that restricts to a nonzero constant on Σ then this process can't keep both A and $A + df$, even though they are two representatives of the same electromagnetic state. So this method of forming the space of configurations that vanish on Σ relies on distinctions present in the representatives that are not present in the physical configurations. And it also distinguishes between gauge-fixed subgroupoids that contain A and those that contain $A + df$. The closest meaningful requirement is that A be gauge-equivalent to 0 on Σ, meaning that there is some smooth function f on Σ such that A restricted to Σ is $-df$. And this requirement is indifferent to any passage to a gauge-fixed subgroupoid.

Respecting equivalences between groupoids may require modifying properties, as we have just seen, and it may also require altering constructions involving multiple spaces. Standard constructions like forming subspaces or forming quotients will unavoidably distinguish between equivalent groupoids, so they require corrections to be well-defined. The general theory of these corrections is extensive. In brief, the category of groupoids

[47] Though the inclusion map in this case would be an equivalence it generally wouldn't be a smooth equivalence. In general there are no subgroupoids with one element per gauge-equivalence class whose inclusion functor is a smooth equivalence. This fact is known as the "Gribov ambiguity", since it was first pointed out by Gribov (1978).

(which we are using to represent the category of spaces of interest) is naturally a homotopical category in the sense of Bill Dwyer, Phil Hirschhorn, Dan Kan, and Jeff Smith (2004). The theory of homotopical categories provides the machinery required to correct any construction that fails to respect equivalences.[48] Only these corrected "homotopical" constructions have any physical significance. I discuss these further in Chapter 2.

The most important role of these homotopical constructions at present concerns the generalization of Yang–Mills configurations to manifolds more general than the whole of Minkowski space. For any manifold M and any group G the groupoid $\Omega^1(M;\mathfrak{g})//G$ of $\mathfrak{g}$-valued 1-forms on M and gauge transformations between these is well-defined. However, if we take this groupoid to be the configuration space of Yang–Mills theory on M then the resulting theory is nonlocal. That is, for two regions U and V of spacetime there are distinct configuration on $U \cup V$ that restrict to the same pair of configurations on U and V, and there are pairs of locally compatible configurations on U and V that are not globally compatible. I discuss this failure of locality extensively in Chapter 3, and I explain there how to formulate a notion of locality that respects equivalence of groupoids. An assignment of configuration spaces to spacetime regions that's local in this way is called a "stack", and any assignment of configuration spaces to regions of spacetime has a "stackification" that results from correcting any nonlocality.[49]

So, finally, in the most generality, I take the configuration space of a Yang–Mills theory with gauge group G on a manifold M to be the stackification of the map sending each region U of M to the groupoid $\Omega^1(U;\mathfrak{g})//G$. This is the description motivated by the considerations of this section: gauge transformations are count-noun identities, and a field theory is local. But the same stack has more familiar descriptions. For example, this stack

[48] The category of groupoids is not just a homotopical category, it is also a left-proper model category and a groupoid-enriched, hence simplicial, model category. These features make homotopy (co)limits particularly easy to compute. See Goerss and Jardine (1999), Hirschhorn (2003), Lurie (2009), Shulman (2009), May and Ponto (2011), and Riehl (2014) for more on these features.

[49] See Chapter 3 for more on stackification, along with Hollander (2008) and Schreiber (2013).

can also be described as map sending each region U to the groupoid of G-principal bundles equipped with a principal connection.[50] And it can be given yet other descriptions in terms of holonomy groups (Rosenstock and Weatherall, 2016). Because these all assign equivalent groupoids to the regions of the spacetime manifolds, they are all different representations of the same space.

1.9 Prospectus

The next three chapters argue for the physical significance of count-noun identities in Yang–Mills theory and general relativity. This introduction has argued that count-noun identities are the appropriate result of the method of symmetry. Adding count-noun identities using the weak quotient removes symmetry, and it removes less structure than the standard way of doing so via the strong quotient. So the method of symmetry only licenses the weak quotient, not the strong. I then sketched a formal treatment of count-noun identities and indicated the kind of physical consequences they can have. The last section gave my interpretation of Yang–Mills theory as an example of a more fully-fledged theory that includes count-noun identities. The following chapters deal with more specific issues in the context of such theories.

Chapter 2 investigates consequences of the homotopical corrections discussed in Section 1.8. In particular, I show that count-noun identities make a difference to the formation of subspaces. Therefore, they affect the cosmological sector of a theory, and the sector of configurations that vanish in some region, and so on. Truncating away identities changes these subspaces, giving another example of disagreement between the

[50] The proof of this requires some piecing together. The functor $\Omega^1(-;\mathfrak{g})//G$ is what Giraud (1971) and Laumon and Moret-Bailly (2000) call a "préchamp" and what Street (1982) calls a "2-separated presheaf". Therefore its stackification can be computed by one application of Grothendieck's plus construction, as they each show. This produces what Kolář et al. call the "physicists version of the connection" (1993, 102), which is also the definition of a connection on a principal bundle given by, e.g., Nicolaescu (1996).

weak and strong quotients. I use this disagreement to respond to Belot's (2018) criticism of the method of symmetry, in which he argues that identifying configurations related by a symmetry produces the wrong sectors. This is only so if one uses the strong quotient to identify configurations; applying the weak quotient and forming homotopically correct sectors gives the right verdict.

Chapter 3 addresses the homotopically correct way of characterizing locality. Intuitively, a theory is local if the state of some region supervenes on the states of that region's subregions. That is, a local theory is one such that there can be no difference in the state of the world without there being some difference in a particular part of the world. So a criterion of locality must appeal to identity and difference of states of the world's parts. In a theory with count-noun identities there is some fact of the matter about which identities are instantiated between the states of the world's parts, and a homotopically correct criterion of locality must account for these. I use this analysis of locality to clarify a point of confusion in the physical and philosophical literatures regarding the locality of different formulations of Yang–Mills theory. Belot (1998) and Richard Healey (2007), among others, have argued for a formulation in terms of properties assigned to curves in spacetime, and debates over the relative merits of this formulation compared to others often contrast the seeming nonlocality of a curve-based formulation with the locality of a field-based one. As I explain in Chapter 3, the locality facts are independent of the differences between a field- and a potiential-based formulation.

Finally, Chapter 4 offers a response to John Earman and John Norton's (1987) "hole argument", which I take to involve two separate issues accompanying count-noun identities. Their first argument, which they call the "verificationist argument", assumes that identity between physical states of affairs must be represented by identity of mathematical models of these states. In Section 1.4 I argued against this assumption, but in Chapter 4 I show how we can take it on board while still keeping count-noun identities. This requires adopting an alternative logic that includes count-noun identities explicitly. The second issue concerns their second argument, the "indeterminism argument", in which they argue that the standard formulation of general relativity is indeterministic. Whether this argument succeeds depends on which of two homotopically acceptable statements of determinism is the correct one. According to one of these statements the standard formulation of general relativity is deterministic. According to the other, which is more in keeping with the spirit of homotopical correctness, the standard formulation of general relativity is indeterministic. Applying the weak quotient to the standard formulation recovers determinism in both senses, and gives a formulation of the theory that more accurately reflects the physics one finds in practice.

2 Sizing up identifications

The more esoteric sections of some Quantum Field Theory (QFT) grimoires speak of two species of gauge transformations, the “large” and the “small”. Small gauge transformations are said to relate two models representing the same physical state, while large gauge transformations are said to relate distinct physical states of affairs that agree on some or all observables—i.e., to be physical symmetries. This size distinction has real, though somewhat arcane, physical consequences: multiple vacua, instantons, and ’t Hooft’s solution of the $U(1)$ problem. But its conceptual status is obscure. It involves invocations of homotopy groups and Euclidean field theory, and interpretive remarks often depend on gauge-dependent features. Indeed, it seems to be a hypothesis independent of the rest of QFT, though various plausibility arguments for it exist.[1] Nevertheless, the distinction has figured in a number of arguments in the philosophical literature. In this paper, I give a derivation of the size distinction that shows it to be a purely kinematical consequence of gauge invariance. In clarifying the conceptual source of the size distinction we also find that it cannot do all of the philosophical work it has been asked to do.

I show that the size distinction results from gauge invariance. Gauge invariance demands that we not physically distinguish gauge-related configurations. Mathematically this is generally accomplished by cooking up some mathematical models of gauge

[1] For textbook treatments of the size distinction and its consequences, see Pokorski (2000, §8.3) and Weinberg (1996, Ch. 23). For a deeper treatment, see Coleman (1985). Finally, see Jackiw and Rebbi (1976) and Callan et al. (1976) for a view from the trenches.

configurations whose isomorphisms are the gauge transformations. Then as long as our reasoning about these models is isomorphism invariant—which is easy to ensure—we will not have violated this demand. But there is a second demand of gauge invariance that this approach does not help us avoid, which is invariance with respect to a choice of gauge fixing. Fixing a gauge is itself isomorphism invariant, so it is harmless. However, when we restrict attention to sectors of a theory—the vacuum or monopole sector of a QFT, for example, or the asymptotically flat sector of general relativity (GR)—it is a subtle matter indeed to state the restriction in a way that does not depend on the choice of gauge fixing. In fact, as I will argue, gauge-invariant membership in one of these sectors is not a property of configurations, but is a *structure* on configurations. As such, the isomorphisms that represent gauge invariance in some sector are the gauge transformations that preserve the membership structure of that sector. In the case of an asymptotically trivial gauge configuration, a small gauge transformation preserves this structure, thus is an isomorphism, thus relates representations of the same physical state. A large gauge transformation is one that does not preserve the structure, so it is not an isomorphism. Thus it is a transformation that relates distinct configurations.

Turning to the philosophical uses of the distinction, I will focus on two arguments: one about intra-theoretical sameness and one about inter-theoretical sameness. Regarding the first, Belot (2018) argues that isomorphisms do not always relate representations of the same physical state, because large gauge transformations are isomorphisms that relate different states. But if my argument is right, large gauge transformations *aren't* isomorphisms, so his counterexamples fail. The second argument arises in the interpretation of Yang–Mills theories. Healey (2007) advances a formulation of these theories for which he claims theoretical advantages, in part because it does not contain the size distinction. This poses a challenge to recent theories of theoretical equivalence that take equivalence of two theories' categories of models to track equivalence of the two theories (Rosenstock et al., 2015). The category of models for Healey's theory is equivalent to that of more traditional formulations, and so ought to come down to the same theory on this accounting. But surely two theories that differ on their theoretical problems cannot be the same theory. As I will argue, Healey's formulation has the same theoretical problems as more traditional formulations, so does not present a counterexample to this picture of theory equivalence.

2.1 The mathematics of the size distinction

I take the size distinction to be an expression of the fact that we must "correct" the equalizer when forming sectors of gauge theories. In this section I explain this cryptic remark. In the next section I apply the corrected equalizer to physical theories to produce the size distinction. Broadly, the idea is this. In gauge theories, the right notion of sameness is not always given mathematically by isomorphism. As such, we must extend our usual mathematical tools. In particular, if we correctly extend the process of forming subspaces to this new context, we find that the isomorphisms within a subspace are not simply the isomorphisms of the entire space that stand between elements of the subspace. So, for example, gauge transformations of an asymptotically vacuum electromagnetic field are not just gauge transformations of that field, but must themselves be asymptotically trivial. Those that aren't—the large gauge transformations—are not isomorphisms in the subspace of asymptotically vacuum electromagnetic fields. The reader content with this gloss can skip to the next section, for the mathematical details are confined to this one.

The relevant formal setting for our discussion of gauge is the theory of groupoids, which has been put to use in the philosophical literature on gauge theory in a number of ways (Rosenstock et al., 2015; Rosenstock and Weatherall, 2016; Weatherall, 2016c). But for motivation we will consider the simpler theory of sets equipped with equivalence relations, or partitioned sets, which form a full subcategory of the category of groupoids.

So we will be interested in the following two categories:

Definition 2.1.1. Let PSet be the category in which an object is a set equipped with an equivalence relation and an arrow $f : X \to Y$ is a function of the underlying sets such that if $x \sim x'$ in X then $f(x) \sim f(x')$ in Y.

Definition 2.1.2. Let Grpd be the category in which an object is a groupoid (a category in which all arrows are isomorphisms) and an arrow is a functor.

These categories are relevantly similar for us because the right notion of sameness in each one is not given by isomorphisms in the sense of structure-preserving maps. In PSet, for example, we have the object $\mathbb{Q}$ whose elements are pairs of integers (a, b) such that b is nonzero and in which $(a, b) \sim (c, d)$ just in case $ad = bc$. We also have an object $\mathbb{Q}'$ whose elements are the rational numbers and in which $q \sim q'$ just in case $q = q'$ for any rational numbers q and q'. Of course $\mathbb{Q}$ is just the representation of the rational numbers using fractions, and we usually write (a, b) as a/b. So $\mathbb{Q}$ and $\mathbb{Q}'$ are just two representations of the same set. But they are not isomorphic in PSet. Any isomorphism in PSet is a bijection on the underlying sets, and the map taking pairs to the rational they represent sends (a, b) and $(2a, 2b)$ to the same rational, so it is not injective. Similar reasoning motivates us to say that isomorphisms in Grpd are not the relevant notion of sameness.

In both of these categories, the right notion of sameness is weaker than isomorphism, so we will call it "weak equivalence". In particular, we have

Definition 2.1.3. An arrow $f : X \to Y$ in PSet is a weak equivalence if the induced function $X/{\sim} \to Y/{\sim}$ on the quotients of X and Y by their equivalence relations is a bijection.

Definition 2.1.4. An arrow $f : X \to Y$ in Grpd is a weak equivalence if it is an equivalence of categories.

In both cases weak equivalence is a strictly wider notion than isomorphism: all isomorphisms are weak equivalences, but some weak equivalences aren't isomorphisms. Returning to our example of the rationals, the natural map sending (a, b) in $\mathbb{Q}$ to the rational it represents in $\mathbb{Q}'$ is a weak equivalence, because (a, b) and $(2a, 2b)$ are sent to the same element of $\mathbb{Q}/\sim$. So in both of these categories the right notion of sameness is not isomorphism. But the usual category-theoretic apparatus was built under the assumption that isomorphism is the right notion of sameness, so we should be prepared for things to go haywire if we press on in preferring weak equivalence.

We will be concerned with the breakdown involved in forming subspaces. Recall that in the category of sets, we form subsets by constructing equalizers. For a pair of set functions

$$X \overset{f}{\underset{g}{\rightrightarrows}} Y$$

the equalizer is the subset of X consisting of elements for which $f(x) = g(x)$. As a special case, if we want to form the subset of X of elements that satisfy some predicate P, then we form the equalizer of the functions

$$X \overset{P}{\underset{\top}{\rightrightarrows}} \{\top, \bot\}$$

where the top function sends any x in X to $\top$ if x satisfies P and $\bot$ otherwise, and the bottom function sends everything to $\top$. In our categories, the equalizer of the above diagram is defined as follows.

Definition 2.1.5. In PSet, the equalizer of f and g is the set of x in X such that $f(x) = g(x)$ equipped with the equivalence relation inherited from X.

Definition 2.1.6. In Grpd, the equalizer of f and g is the groupoid in which an object is an object x of X such that $f(x) = g(x)$, and an arrow $x \to x'$ is an arrow $x \to x'$ in X.

So if we were to apply the usual category-theoretic machinery to the problem of forming subspaces, we would apply these constructions.

The equalizer construction gives the wrong notion of subspaces in PSet and Grpd, however. To see why this is, consider the diagram

$$\mathbb{Z} \underset{(1,1)}{\overset{(-,-)}{\rightrightarrows}} \mathbb{Q}$$

where $\mathbb{Z}$ is the set of integers equipped with the identity equivalence relation, $(-,-)$ is the function that sends some integer n to (n, n), and $(1,1)$ is the function that sends every integer to $(1,1)$. That is, the top map divides an integer by itself, and the bottom map sends every integer to the rational 1. The equalizer of this diagram is then, informally, the collection of integers n such that n/n is 1. Of course, this *should* be all of the integers, since any number divided by itself is 1. But applying the equalizer construction to this diagram actually gives us the set $\{1\}$. This is obviously not what we meant by our description of subspaces, so the equalizer got things wrong. One way to characterize the failure is to note that the equalizer of the diagram

$$\mathbb{Z} \underset{1}{\overset{-/-}{\rightrightarrows}} \mathbb{Q}'$$

gives us the right answer, since $n/n = 1$ for all n. But as we said above $\mathbb{Q}$ and $\mathbb{Q}'$ are just different representations of the same partitioned set, so the subspace-formation operation shouldn't distinguish them. More formally, this characterization says that the equalizer construction is not invariant with respect to weak equivalence, which is the right notion of sameness.

The equalizer construction can be rectified into the homotopy equalizer construction.[2]

[2] For a fuller treatment of the homotopy equalizer in PSet, see Lárusson (2009). For a more general discussion of abstract homotopy theory applicable also to Grpd, see Riehl (2014).

For a diagram

$$X \underset{g}{\overset{f}{\rightrightarrows}} Y$$

in one of our categories, the homotopy equalizer is defined as follows.

Definition 2.1.7. In PSet, the homotopy equalizer of f and g is the set of x in X such that $f(x) \sim g(x)$ in Y, equipped with the equivalence relation inherited from X.

Definition 2.1.8. In Grpd, the homotopy equalizer of f and g is the groupoid in which an object is a pair (x, p) of an object x in X and an arrow $p : f(x) \to g(x)$ in Y, and an arrow $(x, p) \to (x', p')$ is an arrow $q : x \to x'$ in X such that $p' \circ f(q) = g(q) \circ p$ in Y.

In each case, the homotopy equalizer is a rectification of the equalizer in two senses. The homotopy equalizer is a "local" correction to the equalizer in the sense that it is obtained by replacing every reference to equality with the correct notion of sameness in a coherent way (equivalence in the case of PSet and isomorphism in the case of Grpd). It's also a "global" correction because there is a precise sense in which it is the minimal extension of the equalizer from a construction that respects isomorphism to a construction that respects all weak equivalences.

The seed of the size distinction is visible in Def. 2.1.8. Suppose that we have some groupoid C of configurations, and that we would like to restrict attention to some sector of that theory. In this paper we will be especially concerned with the sector of configurations satisfying some boundary condition. There will thus be a groupoid B of boundary conditions and a constant map $b : C \to B$ that sends every configuration to the boundary condition of interest, as well as a restriction map $C \to B$ that sends each configuration to its value on the boundary. The sector of the theory satisfying the boundary condition is then the homotopy equalizer of b and the restriction map. Call this homotopy equalizer E. An isomorphism in E is an isomorphism in C that also preserves the boundary condition, up to isomorphism. Those isomorphisms that don't preserve the boundary condition will still be symmetries in E, since they preserve all of the observables from C—they just won't be isomorphisms, so they will relate distinct configurations. In the case of Yang–Mills theory, an isomorphism in C is a gauge transformation, and an isomorphism in E is a small gauge transformation. So arises the size distinction: large gauge transformations are symmetries, while small gauge transformations relate two models representing the same physical state.

2.2 The physics of the size distinction

Suppose we have a theory that includes a notion of isomorphism of configurations. If we form sectors of our theory in an invariant way, then some isomorphisms of configurations in the whole theory will not be isomorphisms in the sector. In this section, we briefly review some physical consequences of this phenomenon. These examples will be taken from the asymptotically trivial sectors of GR and Yang–Mills theory. In both cases, we find that a global "shift" transformation, like a constant translation in GR, is not an isomorphism. This will be used in the next section to object to Belot's argument that isomorphisms can generate different possibilities. In the case of Yang–Mills theory we find that the theory admits a new fundamental constant θ, which gives rise to a naturalness problem. In the next section I argue that the naturalness problem is formulation-independent, so cannot be solved by adverting to other formulations of Yang–Mills theory.

For simplicity, we restrict attention to theories on the manifold $\mathbb{R}^4$. There are then two configuration spaces of interest:

Definition 2.2.1. In the configuration space of vacuum GR on $\mathbb{R}^4$, a configuration is a Lorentzian metric g on $\mathbb{R}^4$ and an isomorphism of configurations $g \to g'$ is a diffeomorphism[3] $\phi : \mathbb{R}^4 \to \mathbb{R}^4$ such that $\phi_* g = g'$.

[3] We should require further restrictions on ϕ—it should at least be orientation-preserving, for example. Identifying the correct restrictions is a subtle question beyond the scope of this paper. The same goes for the diffeomorphisms of Prop. 2.2.3.

Definition 2.2.2. In the configuration space of a Yang–Mills theory with gauge group G on $\mathbb{M}^4$, a configuration is an $\mathfrak{g}$-valued 1-form A on $\mathbb{M}^4$, with $\mathfrak{g}$ the Lie algebra of G, and an isomorphism $A \to A'$ of configurations is a smooth function $g : \mathbb{M}^4 \to G$ such that $A' = g^{-1}Ag + g^{-1}\,dg$.

In both cases there are multiple isomorphisms between any two configurations, so the discussion of the previous section applies. We construct the asymptotically trivial sectors of each theory as in Def. 2.1.8. A metric in GR is asymptotically trivial if it is Minkowski at infinity, and a gauge field configuration is asymptotically trivial if vanishing at infinity.[4] So the configuration spaces of asymptotically trivial configurations are:

Proposition 2.2.3. In the asymptotically flat sector of vacuum GR on $\mathbb{R}^4$, a configuration is a pair (g, ϕ) of a Lorentzian metric g on $\mathbb{R}^4$ and a diffeomorphism ϕ "at infinity" such that

$$\lim_{r\to\infty} g = \phi^*\eta$$

where η is the Minkowski metric on $\mathbb{R}^4$ and r is the Euclidean distance from the origin. An isomorphism of configurations $(g, \phi) \to (g', \phi')$ is a diffeomorphism $\psi : \mathbb{R}^4 \to \mathbb{R}^4$ such that $\phi_* g = g'$ and

$$\lim_{r\to\infty} \psi = (\phi')^{-1} \circ \phi$$

Proposition 2.2.4. In the asymptotically vanishing sector of a Yang–Mills theory with gauge group G on $\mathbb{M}^4$, a configuration is a pair (A, g) of a $\mathfrak{g}$-valued 1-form A on $\mathbb{M}^4$ and a smooth function $g : \mathbb{S}^3_\infty \to G$ on the sphere at infinity such that

$$\lim_{r\to\infty} A = g^{-1}\,dg$$

[4] The first of these is an oversimplification: there are many notions of asymptotic flatness in GR (Belot, 2018, 35). However, we are only concerned with the qualitative feature that shifts are not isomorphisms in the asymptotically flat sector, and this happens (for at least some shifts) no matter which notion of asymptotic flatness you use. So we choose asymptotically Minkowski for simplicity.

An isomorphism of configurations $(A, g) \to (A', g')$ is a function $h : \mathbb{M}^4 \to G$ such that

$$\lim_{r\to\infty} h = g' g^{-1}$$

From now on we will we will assume that G is a simple Lie group. Since we won't get too far into the details we'll take examples from $G = SU(2)$ for simplicity, though the primary theory of interest is $G = SU(3)$ for the gluons of Quantum Chromodynamics (QCD). Many of the interesting features we encounter in four dimensions are also found in two dimensions with the gauge group $U(1)$, but they are most striking—and most physically relevant—in the four-dimensional case. We now briefly review some physical consequences of moving to the asymptotically trivial sector, with an eye to the arguments of the next section.

First some kinematical differences. The most obvious difference between the full configuration spaces and the asymptotically trivial ones is that some metrics and potentials don't appear in the asymptotically trivial sectors at all. For example, the 1-form

$$A_{\sigma_0} = \sigma_0(x\,dt - t\,dx + z\,dy - y\,dz)$$

with σ_0 the identity in $\mathfrak{su}(2)$ appears in the full configuration space of $SU(2)$ Yang–Mills theory, but not in the asymptotically trivial sector. By excluding these configurations, we are able to define quantities in the asymptotically trivial sector that we could not define in the full configuration space. For any configuration A in the full configuration space we can define the curvature F_A of A. In the asymptotically flat sector, we can then define the quantity

$$c_2(A, g) = -\frac{1}{8\pi^2} \int_{\mathbb{M}^4} \mathrm{tr}(F_A \wedge F_A)$$

c_2 is not well-defined as a function on the full configuration space, because for A_{σ_0} above the integral does not converge. Moreover, it is not isomorphism invariant for those configurations for which it does converge. It is only invariant under isomorphisms corresponding to maps $g : \mathbb{M}^4 \to G$ that are homotopic to the constant identity map. But it is both defined and isomorphism invariant on the asymptotically trivial sector, since a gauge transformation is only an isomorphism in this sector if it is homotopic to the identity.

Less obvious kinematical differences can found in the isomorphism structure of the asymptotically flat sector. Here is one simple difference. Consider the Minkowski metric η, which is naturally an asymptotically trivial configuration (η, id). In the full configuration space, an isomorphism $\eta \to \eta$ is a diffeomorphism $\phi : \mathbb{M}^4 \to \mathbb{M}^4$ such that $\phi_* \eta = \eta$. For example, there is a "temporal shift" diffeomorphism ψ that acts as

$$\psi : (t, x, y, z) \mapsto (t + 1, x, y, z)$$

and we have $\psi_* \eta = \eta$, since ψ is a Poincaré transformation. So ψ is an isomorphism $\eta \to \eta$ in the full configuration space. But it is not an isomorphism $(\eta, \mathsf{id}) \to (\eta, \mathsf{id})$ in the asymptotically trivial configuration space. For this we must have $\psi_* \eta = \eta$ and $\lim_{r \to \infty} \psi = \mathsf{id}$. While the first condition is still satisfied, the latter is not, because ψ is nontrivial at infinity. But it is, of course, still a symmetry in the sense that it preserves any quantities defined only in terms of η. The situation in Yang–Mills theory is similar: a constant gauge transformation is an automorphism of the trivial gauge potential $A_0 = 0$ in the full configuration space but not of (A_0, I) in the asymptotically trivial configuration space, with I the constant identity function. So the asymptotically trivial sectors of these two theories differ from the full configuration space not only in the configurations they contain, but also in their isomorphisms.

Now for the dynamical consequences. In Yang–Mills theory, we find ground state degeneracy. To each configuration (A, g) in the asymptotically flat configuration space we can assign the usual Yang–Mills action:

$$S_{\mathrm{YM}}(A,g) = \int_{\mathbb{M}^4} \frac{1}{2}\mathrm{tr}(F_A \wedge \star F_A)$$

Like c_2, this function is ill-defined on the entire configuration space, because this integral does not generally converge. In the asymptotically trivial sector, for each integer n there is a stationary point (A,g) of the action S_{YM} such that $c_2(A,g) = n$. A solution with $c_2 = 1$ is called an instanton, because it is a localized structure in time, and one with $c_2 = -1$ an anti-instanton. Solutions with other values of c_2 can be approximated by superposing instantons and anti-instantons at different times, splitting the asymptotically trivial configuration space into "instanton sectors", each with their own classical solution. So for each n there is a classical stationary point, and each of these gives a quantum ground state $|n\rangle$. By an appeal to cluster decomposition, this means that the physical vacuum is a coherent superposition (Weinberg, 1996, §23.6, Callan et al., 1976, 2721):

$$|\theta\rangle = \sum_{n=-\infty}^{\infty} e^{in\theta}|n\rangle$$

In this expression θ is a new fundamental constant. It has empirical consequences when matter is introduced, and in the case of QCD observation has shown that $|\theta| < 10^{-9}$. This is a naturalness problem, made all the more pressing by the fact that if $\theta \neq 0$ then the strong force violates CP symmetry. As such, it goes by the name "the strong CP problem".

Before addressing our philosophical topics in the final section, a word on how my derivation of the size distinction relates to previous arguments for it. Previous arguments for the size distinction fall into two classes: arguments about the well-definition of various physical quantities, and analyses in the constrained Hamiltonian framework. The former originate in the early days of the size distinction in gauge theories, and can also be found in the literature on GR. In the gauge case, Jackiw (1980, 666) argues that if we allow large gauge transformations to be isomorphisms, then the total Noether charge corresponding to the gauge symmetry will be ill-defined, because not preserved by isomorphism. This is a consequence of the discussion of this section: the total charge coincides with c_2, and as we remarked above this quantity is only invariant under the isomorphisms of the asymptotically trivial sector—i.e., the small gauge transformations.

A typical argument in the second class is concerned with Gauss's Law in the constrained Hamiltonian formulation of gauge theories. Balachandran (1994, 21), for example, argues that in order for the Poisson bracket between Gauss's Law and physical quantities to be well-defined we must impose some asymptotic conditions on gauge transformations, and that the simplest condition is asymptotic triviality. Giulini (1995), too, argues that in GR and gauge theories we ought to take large gauge transformations (or large diffeomorphisms) to relate representations of different physical situations, because they do not respect the Gauss's Law constraint. As with the first class of argument, we can see this distinction as flowing from the size distinction as I presented it: the Gauss's Law constraint is a manifestation of Stokes's theorem combined with the fact that c_2 is well-defined. It appears as a constraint when one picks a preferred gauge, as standard treatments of constrained Hamiltonian dynamics do.

My derivation of the size distinction improves on these arguments in two ways. First, it does not require any assumptions about dynamics or which quantities are well-defined; these are, rather, consequences of the distinction. These assumptions significantly weaken philosophical applications of the size distinction. In GR in particular the constrained Hamiltonian formulation has been the subject of substantial skepticism (Maudlin, 2002). But in the Yang–Mills case, too, reliance on Hamiltonian dynamics makes the size distinction "sensitive to theoretical context" (Healey, 2007, 183) in a potentially worrisome way. Second, this derivation of the size distinction shows that it is an instance of a much more general phenomenon. Any time that we consider sectors of a theory in which there are isomorphisms between configurations we should expect to find similar results. This gives us a systematic response to arguments like Belot's, which we now consider.

2.3 The philosophy of the size distinction

In the previous sections we saw that treating sector formation in an invariant way can change both the elements of a configuration space and what counts as an isomorphism. This has physical consequences, including the revelation of a new fundamental constant. In this section we consider two philosophical consequences. First, it allows us to maintain that isomorphisms cannot relate distinct physical states of affairs. Belot (2018) argues that isomorphism comes apart from sameness of state by showing that physicists' sameness judgements vary depending on which sector of the theory they are investigating. But if the story above is correct, then this is exactly what we should expect. Our second consequence concerns the picture of theory equivalence according to which a theory is a category of models. Healey (2007) claims that the strong CP problem, encountered in the previous section, does not arise in his formulation of Yang–Mills theory, giving it an advantage on potential-based formulations. But we saw in the previous section that little hangs on the use of potentials; the strong CP problem comes about from vacuum degeneracy, and this because the size distinction partitions the asymptotically trivial configuration space into instanton sectors. Since all of this is (constructed to be) invariant under equivalence of categories, the strong CP problem arises in a formulation like Healey's just as much as it does in a field-based formulation.

Belot (2018) is concerned with the question of when isomorphisms can generate new possibilities. There are two popular answers to this question among philosophers. According to the old, "shifty" account, isomorphisms generate new possibilities just in case they are "shift" isomorphisms. In spacetime physics these are diffeomorphisms like the one of the previous section, which send every point of spacetime one unit to the future (or the past, or the right or the left). In gauge theories, shifts are the gauge transformations that take on the same value at every point in spacetime. These are to be contrasted with generalized shifts, which in spacetime physics are arbitrary diffeomorphisms, an in gauge theory arbitrary gauge transformations. In the full configuration space of the theory, both shifts and generalized shifts are isomorphisms. But for shifty philosopher shifts relate distinct states of affairs and generalized shifts do not. The second, newer, "shiftless" account is a near consensus among philosophers of physics (Belot, 2018, fn. 27). On this view, isomorphisms can never relate distinct physical states of affairs, whether they be global shifts or generalized. The shiftless and the shifty only disagree about global shifts; when it comes to generalized shifts both sides agree that they relate the same state of affairs.

According to Belot (2018), both sides are wrong to say that generalized shifts never relate distinct states of affairs. As evidence he proffers a bevy of apparent counterexamples from physicists: configurations that are isomorphic but which physicists take to be distinct. Consider Belot's central example: asymptotically trivial isomorphisms. Belot argues that in some sectors of GR isomorphic configurations are nonetheless physically distinct. I think his arguments work just as well for Yang–Mills theories. And the lesson of these arguments is this: "while relativists *do* often speak as if solutions of general relativity are gauge equivalent if and only if [isomorphic], they drop this way of speaking when asymptotic boundary conditions... are in view" (2018, 41). So it looks like both the shifty and the shiftless are wrong to say that generalized shifts never generate new physical states of affairs. If a generalized shift is asymptotically nontrivial it does just this.

The discussion of the previous two sections came to the same verdict as Belot. But it did so by invoking shiftlessness. Above, we claimed that small gauge transformations relate the same state of affairs because they are isomorphisms, and large gauge transformations relate distinct states of affairs because they are not isomorphisms. Taking isomorphism and sameness of state to coincide like this is just shiftlessness. So rather than being a counterexample, the size distinction shows that shiftlessness gets the cases *right*.

This said, the shiftless picture of the previous two sections is not the canonical forms of shiftlessness that Belot criticizes. Indeed, the picture above is an implementation of some of Belot's conclusions. The major difference between our picture and the usual shiftlessness is, of course, the way in which isomorphisms in the full theory relate to isomorphisms of sectors of the theory: the transformations they induce in a sector are only isomorphisms if they also preserve the structure of belonging to that sector. But this difference follows from the more general assumption that we ought to treat configuration spaces as being equipped with isomorphisms, not as mere sets. And this is one of Belot's conclusions: "in philosophy of physics, we should spend more time thinking about spaces of solutions—with the structures and group actions that they carry" (2018, 50). So while I take shiftlessness to ultimately win out, it is a somewhat generalized form of shiftlessness.

As a consequence of this generalization we see new ambiguities in old theories. In the case of Yang–Mills theories, for example, I argue in Chapter 3 that there is an important difference between the configuration space described in Def. 2.2.2—in which we equip the set of gauge potentials with isomorphisms representing gauge transformations—and the set of gauge-equivalence classes of potentials. As he argues, the former is local in a way that the latter is not. The size distinction is another difference between these spaces. If the full configuration space of Yang–Mills theory is the set of gauge-equivalence classes of potentials, then applying Def. 2.1.8 gives the set of gauge-equivalence classes of potentials asymptotically gauge-equivalent to the zero potential as the configuration space of the asymptotically trivial sector of the theory. And in this theory there is no size distinction, because isomorphism coincides with equality of equivalence classes. The sector-dependence of isomorphism that we need to get the size distinction is only possible if we keep specific isomorphisms around, not if we collapse everything into isomorphism classes.

Keeping track of isomorphisms like this allows us to solve the puzzle about the strong CP problem. Healey (2007, §7.5) argues that if we instead take the configuration space of Yang–Mills theory to be given by holonomy maps—assignments of group elements to curves in spacetime—then the strong CP problem mentioned above does not arise. This would be a theoretical advantage of such a formulation, because a naturalness problem would be eliminated. According to Healey (following Fort and Gambini, 2000), "when a theory is formulated in a loop/path representation, all states and variables are automatically invariant under both "small" and "large" gauge transformations, so there is no possibility of introducing a parameter θ" to parametrize the superposition involved in the physical vacuum $|\theta\rangle$ above (2007, 198). But this creates a puzzle for a certain increasingly popular picture of theory equivalence, according to which two theories are equivalent if their categories of models are equivalent as categories. As Rosenstock and Weatherall (2016) show, curve-based formulations like Healey's give categories of models that are equivalent to the category described in Def. 2.2.2. So according to this picture of theory equivalence, formulations like Healey's give the same theory as potential-based formulations like Def. 2.2.2. But this conflicts with the claim that curve-based formulations have the theoretical advantage of dissolving the strong CP problem, for surely one theory is *better* than the other if it solves more problems, not the same.

The solution to this puzzle is, again, keeping better track of isomorphisms. As I point out in Chapter 3, there are two classes of curve-based formulations of Yang–Mills theory. In the first class, one takes the configuration space to be given by the set of isomorphism classes of holonomy maps. In this configuration space it's true that there is no size distinction, so the strong CP problem cannot arise. But this configuration space is equivalent to the set of gauge-equivalence classes of gauge potentials, and there, too, as we just saw, there is no size distinction either. On the other hand, we might take the configuration space to have holonomy maps as objects and holonomy isomorphisms as isomorphisms. In this case, as

Rosenstock and Weatherall show, the configuration space is equivalent to the configuration space described in Def. 2.2.2. And the sector-formation construction from Def. 2.1.8 is designed to respect equivalences like this, so the asymptotically trivial sector of Yang–Mills theory described in Prop. 2.2.4 is equivalent to the asymptotically trivial sector of the theory formulated in terms of curves. And in this case the curve-based formulation has the size distinction after all: small gauge transformations are isomorphisms, large gauge transformations are not, and each instanton sector of the theory has its own classical solution to the Yang–Mills equation.

2.4 Conclusion

I have advanced three related arguments. First: the size distinction in Yang–Mills theory and GR is a consequence of forming sectors of theory in an isomorphism-invariant way using the homotopy equalizer. If take our configuration spaces to be given by sets of configurations equipped with isomorphisms between them, then we find that restricting attention to sectors of our theory also requires restricting isomorphisms to those which preserve the sector membership structure. In the case of Yang–Mills theories, these are just the small gauge transformations, so the remaining large gauge transformations relate distinct states of affairs.

The other two arguments involved interpretive questions about sameness in the philosophy of physics. Appealing to phenomena like the size distinction, Belot (2018) argues that isomorphism must come apart from sameness of physical state. But this is only true if you treat sector formation naively, rather than as we have here. Indeed, the above approach explicitly relies on the claim that isomorphism and sameness coincide to explain why the size distinction arises. The physical consequences of the size distinction lead to a certain theoretical puzzle, and Healey (2007) argues that an alternative formulation of

Yang–Mills theory avoids this puzzle by avoiding the size distinction. But if we are careful about the source of the puzzle and the statements of the various formulations, we find that the size distinction and the puzzle it causes are independent of how the theory is formulated.

Chapter 2, in full, is a reprint of material currently being prepared for submission for publication. The dissertation author is the sole author of this material.

3 Identifications and separability

Much of the philosophical literature on Yang–Mills theories is concerned with the differences between interpretations that represent the state of the world using fields and those that represent the state of the world using properties assigned to curves in spacetime. These interpretations are inspired by corresponding mathematical formulations of Yang–Mills theory: using bundles and holonomies, respectively. The former class—advocated by Weatherall (2016a), for example—is generally thought to deliver a localized picture of the world, but also to involve a kind of "surplus structure" (Redhead, 2001). On these interpretations, mathematically unequal but gauge-equivalent configurations correspond to the same physical state of affairs, so there is a representational redundancy. The latter class of interpretations—advocated by Belot (1998) and especially Healey (2007), for example—is meant to eliminate this surplus structure at the cost of locality. In these theories, the state of some region does not supervene on the state of its subregions (Myrvold, 2011). The interpretive choice between these positions is sometimes presented as a cost-benefit analysis, a trade-off between locality and surplus structure.

This accounting oversimplifies the situation. Every claim above is contested: Lyre (2004) and Wallace (2014) argue that field-theoretic interpretations are nonlocal, and Rosenstock and Weatherall (2016) argue that the two classes of interpretations have the same amount of structure. This calls into question the advantages of each interpretation. If bundle interpretations are nonlocal, then they have no advantages over holonomy interpretations and have unnecessary structure on top of that. But if holonomy interpretations have the same amount of structure, then they have no less structure than bundle representations, and so have whatever surplus the bundle interpretations do, too. Sorting out what's really going on here is made difficult by the fact that, by and large, arguments about locality in this literature are semiformal at best. We should like a precise statement, susceptible to proof, of the locality facts in these theories.

I show that the disagreements over locality and structure result from equivocation. There are two classes each of bundle representations and holonomy representations that differ in their structure and locality facts. One class of representations, call them truncated, has less structure than the other class of untruncated representations. According to the definition of locality given below, truncated Yang–Mills representations are nonlocal while untruncated representations are local. Indeed, the part of the representation that gets lopped off by truncation is just the structure representing the locality of the theory. So the distinction between local and nonlocal theories is orthogonal to the distinction between bundle and holonomy representations: each kind of representation has a local and a nonlocal version. The usual story attributes locality to bundle representations and nonlocality to holonomy representations because it considers only the untruncated bundle representations and only the truncated holonomy representations. When we make precise what we mean by locality, it becomes clear that the relevant feature of the theory is whether it is truncated, not whether it involves bundles or holonomies.

Disambiguating these theories also leads to an interesting general lesson. On the standard view, a physical theory is its set of models. Halvorson (2012) has argued that this conception of theories does not do justice to facts about relationships between theories: it identifies distinct theories and distinguishes alternative formulations of the same theory. It also fails to capture ways that one theory might be a specialization or generalization of another. Halvorson concludes from this that a physical theory cannot solely be its collection of models; instead we must keep track of isomorphisms between these models, in the style of Rosenstock et al. (2015). In what follows, we find that these "external" facts about relationships between theories are not the only thing that the standard view misses out on. Local and nonlocal gauge theories have the same set of models, but cannot be the same theory precisely because one is local and the other is not. Articulating the difference between these theories requires appealing to other parts of the mathematical structure—in particular, the sameness structure of the theory. Truncating a theory forgets precisely this sameness structure, simultaneously rendering the theory nonlocal. So a view that takes a physical theory to be its set of models also fails to account for "internal" facts about the theory itself.

3.1 Spaces of gauge configurations

I claimed that "bundle formulation" and "holonomy formulation" are importantly ambiguous. In this section we resolve this ambiguity before considering locality. Briefly, we must choose between modeling gauge-related configurations as equal and modeling them as isomorphic. If we go the former route, the theory is nonlocal whether it is a bundle formulation or a holonomy formulation. If we instead model gauge equivalences as isomorphisms, either kind of formulation is local.

Let X be some spacetime region, and consider electromagnetism as a $U(1)$ Yang–Mills theory. What is the configuration space of the gauge fields on X? According to the standard story, there are two classes of answers to this question (Belot, 1998; Healey, 2007). The first class takes the bundle formulation of Yang–Mills theory as its point of departure. According to this approach, a configuration is specified by a connection on a principal $U(1)$-bundle or by some coordinate representation thereof like a vector potential. The second class of answers takes the configuration space to be a collection of holonomy maps, which assign elements of $U(1)$ to curves in X. Many debates in the philosophical literature are concerned with this level of generality, and arguments for or against one class of formulations confer support to all members of that class.

But as I've presented these formulations, they are incomplete. I have only said how to specify a configuration, not how to determine whether two configurations are the same. Specifying the space of gauge configurations requires choosing when two mathematical representatives correspond to the same physical state of affairs. It is generally thought that there are two choices in the case of Yang–Mills theory. This is incorrect. To clearly lay out the correct menu of choices in a Yang–Mills theory, it helps to speak in terms of groupoids, which make these sameness facts explicit.[1]

Definition 3.1.1. A *groupoid* is a category in which all arrows are isomorphisms.

Much of the philosophical literature on the interpretation of Yang–Mills theories already involves groupoids. For example, Rosenstock and Weatherall (2016) argue that bundle and holonomy interpretations have the same amount of structure, in the sense that a particular groupoid of bundles is equivalent to a particular groupoid of holonomy maps. Weatherall (2016c), too, uses groupoids to analyze relative amounts of structure in gauge theories. More generally, Rosenstock et al. (2015) adopt a point of view on which a scientific theory amounts to a groupoid of models, rather than the set of models considered on the standard semantic view of theories.

I will assume for the rest of this discussion that every collection of mathematical objects is a groupoid. That is, I will take them to generalize sets. A description of a collection requires both a description of the objects in the collection and a description of when two objects are the same—i.e., of objects and isomorphisms. In some cases the isomorphism structure is trivial; these are the cases corresponding to sets. More precisely,

[1] For an introduction to groupoids, see Brown (1987).

say that a groupoid is a set if there is at most one isomorphism between any two of its objects.

Most discussions in the philosophical literature assume that any collection must be a set. So when these discussions encounter a groupoid—such as the groupoid of principal bundles and principal bundle isomorphisms between them—they must squash the isomorphism structure into triviality to produce a set. Call this process truncation.

Definition 3.1.2. For any groupoid X, its *truncation* $\|X\|$ is the set of isomorphism classes of X, considered as a groupoid.

If a groupoid X is a set (in the sense of the previous paragraph), then it is equivalent to its truncation $\|X\|$. If it is not a set, then there is a natural map $X \to \|X\|$ that sends each object to its isomorphism class, but this map has no inverse. So truncation forgets information about a groupoid, in particular information about its isomorphism structure.

Return to the candidate configuration spaces for a Yang–Mills theory. For simplicity, consider a simply-connected spacetime region X, and consider a formulation in terms of vector potentials. Let $\Omega^1(X)$ denote the set of all such vector potentials. If we take this set to be the configuration space, then we have a well-known problem: any two vector potentials related by a gauge transformation are empirically and dynamically indistinguishable, leading to empirical underdetermination and dynamical indeterminism. In response, physicists take gauge-related potentials to represent the same physical configuration. That is, the set $\Omega^1(X)$ is too fine, because it includes spurious distinctions between gauge-related vector potentials.

We cannot take the set of vector potentials to be the configuration space. Instead we want a space that is obtained from $\Omega^1(X)$ by identifying any two gauge-related configurations. There are two ways to do this, using groupoids. If we restrict ourselves to sets, there's only one option: take the set of gauge-equivalence classes of potentials. But groupoids allow for non-trivial sameness structure, and we can take advantage of this. Let $\Omega^1(X)//U(1)$

be the groupoid in which an object is a vector potential and an isomorphism is a gauge transformation. This is a kind of quotient of $\Omega^1(X)$, in that it starts with the set of potentials and then makes any gauge-related configurations the same. But it is not the familiar set-theoretic quotient into equivalence classes. That, rather, is the set $\|\Omega^1(X)//U(1)\|$, the truncation of our weaker quotient.

Assume that gauge-related potentials are the same physical configuration. It remains to decide whether the configuration space is the groupoid where gauge transformations are isomorphisms or the truncation thereof. Relaxing the assumption that our spacetime is simply connected, one presentation of the bundle and holonomy formulations of the configuration space is the following:[2]

Definition 3.1.3. Let X be a smooth manifold and G a Lie group with Lie algebra $\mathfrak{g}$. Denote the *groupoid of principal G-bundles with connection on X* by $\mathbf{B}G_{\text{conn}}(X)$, in which an object is a principal G-bundle on X equipped with a principal G-connection, and an isomorphism is a principal bundle isomorphism that restricts to the identity on the base space and preserves the connection.

Definition 3.1.4. Let X be a smooth manifold and G a Lie group. Denote the *groupoid of G-valued holonomy maps on X* by $[\mathbf{P}_1(X), \mathbf{B}G]$, in which an object is a map from the path groupoid of X to the group G considered as a one-object groupoid, and an isomorphism is an isomorphism of such functors.

Both of these definitions are underspecified, but we are concerned with a high-level distinction between bundle and holonomy formulations, so this is fine. It will suffice for our discussion to keep in mind the simpler electromagnetic case of the groupoid $\Omega^1(X)//U(1)$, where an object is a vector potential and an isomorphism is a gauge transformation. When X is simply connected, this groupoid is equivalent to $\mathbf{B}G_{\text{conn}}(X)$.

[2]These definitions are taken from Schreiber (2013, §1.2.6.1.1); see there for more rigorous statements.

Neither of the groupoids just defined is a set, and so neither is equivalent to its truncation. So we seem to have four candidate configuration spaces: the bundle formulation $\mathbf{B}G_{\text{conn}}(X)$, the holonomy formulation $[\mathbf{P}_1(X), \mathbf{B}G]$, and the truncations of each. However, it follows from Rosenstock and Weatherall's (2016) main result that $\mathbf{B}G_{\text{conn}}(X)$ and $[\mathbf{P}_1(X), \mathbf{B}G]$ are equivalent, so they are two presentations of the same groupoid. Hence their truncations are also equivalent. As such, there are really only two options: the groupoid $\mathbf{B}G_{\text{conn}}(X)$ and its truncation $\|\mathbf{B}G_{\text{conn}}(X)\|$. And these really are different, because the former is not a set and the latter is.

This menu of spaces resolves one of the tensions discussed in the opening. The standard story has it that bundles involve more structure than holonomy models, but Rosenstock and Weatherall show that there is a precise sense in which a particular bundle representation has the same amount of structure as a particular holonomy representation. As we now see, the standard story rests on an ambiguity. By the measure of structure Rosenstock and Weatherall use, $\mathbf{B}G_{\text{conn}}(X)$ and $[\mathbf{P}_1(X), \mathbf{B}G]$ have the same amount of structure, and their truncations have the same amount of structure, but $\mathbf{B}G_{\text{conn}}(X)$ and $[\mathbf{P}_1(X), \mathbf{B}G]$ have more structure than their truncations. In the standard story, "the" bundle representation is usually taken to be $\mathbf{B}G_{\text{conn}}(X)$, and "the" holonomy representation is usually taken to be $\|[\mathbf{P}_1(X), \mathbf{B}G]\|$. So, properly interpreted, this story is right. It's just not the whole story. Worse, this terminology has not be consistently maintained. Wu and Yang (1975), for example, do not distinguish between $[\mathbf{P}_1(X), \mathbf{B}G]$ and its truncation, referring to the truncated state space as merely "less easy to use" (1975, 3846) than the untruncated one. But the details of their argument rely on structure lost in truncation, so this is a matter of substance, not convenience. A defense of a holonomy interpretation based on the truncated state space's having less structure must show that the truncated state space has all the resources Yang–Mills theory requires, so it cannot appeal to an analysis like Wu and Yang's, which does not attend to the distinction.

Turn to the second tension, concerning locality.

3.2 Locality, functorially

The particular notion of locality that interests us is separability. A physical theory is separable if the physical state of some region supervenes on the physical states of its subregions. To make this precise, we need a way of spelling out how the mereological structure of spacetime is reflected in the structure of the space of possible states of regions. For any field theory, the state of a region involves at least the states of its subregions. In particular, there is a duality between parthood and determination: if U is a subregion of X, then a physical state of X induces a physical state of U when we restrict our attention. The assignment of configuration spaces to spacetime regions is thus functorial—it respects the composition structure of spacetime. If the theory is separable, then the state of a region also involves no more than the states of its subregions. So we will formalize separability as a property of the functor that assigns configuration spaces to regions of spacetime. Informally, this functor is separable if the configuration space it assigns to some region is the same as the space of collections of states of its subregions.

Einstein formulated the earliest version of the principle of separability as an articulation of a difference between classical and quantum theories. Broadly, a theory is separable if the state it assigns to a system is determined by the state assigned to its subsystems. For a classical field theory, these systems are regions of spacetime, with subregions as subsystems. Einstein's principle can be sharpened up into a more formal criterion, versions of which have been given by Belot (1998, 544), Healey (2007, 125), and Myrvold (2011). Following Myrvold, take some manifold X and consider covers of X by open sets. A cover $\mathfrak{V}$ of X is finer than a cover $\mathfrak{U}$ if every region in $\mathfrak{V}$ is a subregion of some region in $\mathfrak{U}$. We

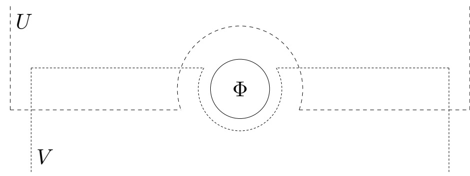

Figure 3.1: The Aharonov–Bohm setup

can give a semiformal definition of separability as follows:[3]

> For any spacetime region X, there are arbitrarily fine open covers $\mathfrak{U}$ of X such that the state of X supervenes on the states of the elements of $\mathfrak{U}$.

As these authors have pointed out, the Aharonov–Bohm (AB) effect (Aharonov and Bohm, 1959) demonstrates a failure of separability for holonomy formulations. If $\{U, V\}$ is a cover of the exterior region of an infinite solenoid containing magnetic flux Φ by simply-connected regions, as in Fig. 3.1, then there is only one possible state for each of U and V, represented by the trivial holonomy. But the state of $U \cup V$ depends on the current in the solenoid, and there are infinitely many different ways that $U \cup V$ could be. So a difference in the state $U \cup V$ does not imply a difference in the states of U or V.

However, this criterion is still ambiguous. It formalizes the mereological structure of spacetime, but the notion of supervenience remains intuitive. To formalize the claim that the states of the subregions determine the states of the entire region, we need to model the way that possibilities for subregions interact with the mereological structure of spacetime. For a field theory like Yang–Mills theory, the state of a region determines the state of its subregions: if you tell me the electromagnetic facts in this building, I can tell you the electromagnetic facts in each room. More formally, a Yang–Mills theory has a presheaf of configuration spaces:[4]

[3]This analysis is adapted from (Myrvold, 2011, 425). See also Healey (2007, 46) and Belot (1998, 540).
[4]See Mac Lane and Moerdijk (1992) for an introductory treatment of presheaves of sets, which are a special case of our presheaves.

Definition 3.2.1. Let X be a topological space. A *presheaf (of groupoids) on* X is a functor $F : \mathcal{O}(X)^{\mathsf{op}} \to \mathsf{Grpd}$, where Grpd is the category of groupoids and $\mathcal{O}(X)$ is the category in which an object is an open set $U \subseteq X$ and for any objects V and U there is a unique arrow $V \to U$ just in case $V \subseteq U$.

Presheaves capture the duality remarked upon at the start of this section. If V is a subregion of U, then there is an arrow $V \to U$, and a presheaf F sends this to an arrow $F(U) \to F(V)$. If we think of F as an assignment of possibility spaces to regions of spacetime, then this map $F(U) \to F(V)$ is just the restriction map.

In physics, a field theory involves a presheaf of configuration spaces on spacetime. For example, consider a theory involving some scalar field on a spacetime X—a mass density, say, or a gravitational potential. The possible field configurations on X are elements of the set $C^\infty(X, \mathbb{R})$ of real-valued functions on X. If U is a subregion of X, then the possible configurations of U are elements of the set $C^\infty(U, \mathbb{R})$. Given a configuration ϕ on X in $C^\infty(X, \mathbb{R})$, we obtain a configuration $\phi|_U$ in $C^\infty(U, \mathbb{R})$ via restriction. So the presheaf of configurations in this theory is the functor $C^\infty(-, \mathbb{R}) : \mathcal{O}(X)^{\mathsf{op}} \to \mathsf{Set}$.

If we think of separability in terms of presheaves, it amounts to the claim that the restriction maps can sometimes be reversed. Consider the AB setup. By restriction, a configuration in $F(U \cup V)$ gives a configuration in $F(U)$ and one in $F(V)$, and these configurations agree on $F(U \cap V)$. So there is a map taking a configuration in $F(U \cup V)$ to a collection of compatible configurations on the cover. If the theory is separable, we can go the other way: a collection of compatible configurations on the cover determines a configuration in $F(U \cup V)$. Moreover, these two maps are inverses. Two different elements of $F(U \cup V)$ are sent to different collections of configurations on the cover, because there can be no difference in the state of $U \cup V$ without a difference in the state of some subregion in the cover—this is just what it means for the state of $U \cup V$ to supervene on the state of its subregions. This makes the restriction map injective, and since we have a determination map in the other direction it must be a bijection.

Formalizing the compatibility condition on a collection takes a bit of care if it is to be done in a gauge-invariant way. To see how things could go wrong, consider the presheaf of configuration spaces $\Omega^1(-)//U(1)$. Naively, a compatible collection of configurations for the AB setup is a pair of vector potentials A_U on U and A_V on V such that $A_U = A_V$ on the overlap. However, this statement of the compatibility condition is not gauge-invariant. If (A_U, A_V) is such a pair (for example, $A_U = A_V = 0$) and A'_V is gauge-equivalent to A_V (for example, $A'_V = d\theta/2\pi r$), then (A_U, A'_V) will generally not be a compatible collection. But (A_U, A_V) and (A_U, A'_V) are gauge-equivalent, because A_V is gauge-equivalent to A'_V. So this statement of the compatibility condition distinguishes between isomorphic collections. This makes it gauge-variant and ill-defined with respect to the groupoid structure. So this cannot be the right statement of the compatibility condition.

The problem we face is familiar to mathematicians, and they have developed a extensive toolkit called abstract homotopy theory to deal with problems like it. In our AB case, the theory says that the groupoid of compatible collections of configurations on the cover is the groupoid where an object is a triple (A_U, A_V, g) of configurations A_U and A_V on U and V, respectively, and a gauge transformation g from A_U to A_V on the overlap $U \cap V$. An isomorphism in this groupoid between objects (A_U, A_V, g) and (A'_U, A'_V, g') is a pair of gauge transformations h and h' between the first two entries of the triple such that $hg' = gh'$.

Generalizing this situation, for any presheaf F on a space X and cover $\mathfrak{U}$ of X we define the collection of compatible configurations using the homotopy limit, which turns a diagram of configuration spaces and restriction maps into a groupoid of collections of configurations that are compatible with respect to all of the restriction maps in an isomorphism-invariant way. If the configuration space $F(X)$ is the same as the space of compatible collections for every cover, then we say that F is separable, since it is separable with respect to every cover. In mathematical parlance this means that F is a stack.[5]

Definition 3.2.2 (Hollander, 2008, Def. 1.3)**.** For any topological space X, a presheaf F on X is a *stack on* X if for any good cover $\mathfrak{U}$ of X the natural map

$$F(X) \to \operatorname{holim}\left(\prod_{U \in \mathfrak{U}} F(U) \rightrightarrows \prod_{U,U' \in \mathfrak{U}} F(U \cap U') \mathrel{\substack{\to\\\to\\\to}} \cdots\right)$$

is an equivalence. Call this homotopy limit the *groupoid of descent data* for F with respect to $\mathfrak{U}$.

Note that the stack condition is well-defined, in that it does not distinguish between equivalent presheaves: if two presheaves are equivalent and one is a stack, then so is the other. Informally, Def. 3.2.2 says that a presheaf is a stack if a configuration of the region X is the same thing as a compatible collection of configurations of the subregions of X, for any way of carving X up into subregions.

I take the stack condition to be a precise statement of separability. For the concept to apply, the theory must assign configuration spaces to all of the subregions of some spacetime region, and the state of the region determines the state of its subregions. So we should be concerned with presheaves. The stack condition then formalizes the idea that the state of the entire region supervenes on the states of its subregions. Any two different collections of compatible subregions determine different configurations of the total region, and vice versa. So there cannot be a difference in the configuration of the total region without a difference in some subregion. With this analysis of separability in hand, we can ask whether our Yang–Mills theories are separable.

[5]The requirement that the cover be "good" is a technical condition encoding the "sufficiently fine" clause in the semiformal statement above. See Schreiber (2013) for a discussion of its importance.

3.3 Separability in gauge theories

The untruncated groupoids of Defs. 3.1.3 and 3.1.4 generalize to stacks; their truncated cousins do not. In closing, I will sketch proofs of these claims. The arguments from Lyre (2004) and Wallace (2014) are proofs of the latter fact. Myrvold's (2011) proof of the nonseparability of Healey's (2007) holonomy theory is dual to these, showing that the pre-cosheaf assigning algebras of observables to spacetime regions is not a costack. Since observables are, broadly speaking, functions on state space, this argument is dual to Lyre's and Wallace's. Benini et al. (2015) give a similarly dual argument that the untruncated algebras of observables are a costack. The fact that the configuration space of a gauge theory is a stack is urged by Schreiber (2013), and the discussion here broadly follows his.

Consider one last time the AB setup. We can take the configuration in the exterior region to be (the principal connection corresponding to) the potential $A = \Phi/2\pi r\, d\theta$, where Φ is the magnetic flux through the solenoid. Suppose that A and A' are the potentials corresponding to fluxes Φ and Φ', respectively. A and A' are gauge-equivalent just in case Φ and Φ' differ by a multiple of 2π; otherwise they specify two different equivalence classes $[A]$ and $[A']$ in the configuration groupoid $\|\mathbf{B}U(1)_{\text{conn}}(U \cup V)\|$. The groupoid of descent data in this case is the set of pairs of gauge-equivalence classes of configurations on U and V which are equal on the overlap. But no matter the value of Φ, the restriction $A|_U$ to U is gauge-equivalent to the vanishing potential, and likewise for $A'|_U$. It follows that both $[A]$ and $[A']$ are mapped to the same descent datum, the pair $([0],[0])$. Since generally $[A] \neq [A']$, the map appearing in the stack condition is not injective. Hence it is not an equivalence, and we have shown that

Proposition 3.3.1. The presheaf $\|\mathbf{B}G_{\text{conn}}(-)\|$ is generally not a stack.

It follows from this proposition and the equivalence between $\mathbf{B}G_{\text{conn}}$ and $[\mathbf{P}_1(-), \mathbf{B}G]$ that the presheaf $\|[\mathbf{P}_1(-), \mathbf{B}G]\|$ is also not a stack. So both of the truncated configuration spaces are nonseparable.

The untruncated configuration spaces avoid this counterexample, and we can prove that they satisfy the stack condition. The configurations in $\mathbf{B}G_{\text{conn}}(U \cup V)$ corresponding to A and A' are not isomorphic unless the difference in their respective fluxes Φ and Φ' is a multiple of 2π. The former is mapped to the descent datum $(A|_U, A|_V, 1)$, and the latter to $(A'|_U, A'|_V, 1)$, where the third entry in both is the gauge transformation corresponding to the constant identity function on $U \cap V$. As before, we can use the fact that both potentials are gauge-equivalent to 0 when restricted to U or V to transform them away. The triple corresponding to A becomes $(0, 0, e^{i\lambda})$ and the triple corresponding to A' becomes $(0, 0, e^{i\lambda'})$, where λ is a locally constant real-valued function on $U \cap V$ such that the difference $\lambda|_U - \lambda|_V$ is equal to Φ, and likewise for λ' and Φ'. So the two triples are not the same: the third entry of the triple retains the information about the magnetic flux up to a multiple of 2π. The isomorphism structure of the groupoid allows the compatibility condition to encode the difference between the two configurations, avoiding the counterexample.

More generally, we have the following

Proposition 3.3.2. $\mathbf{B}G_{\text{conn}}$ is a stack.

Proof sketch. Let $\mathfrak{g}$ be the Lie algebra of G, and let $\Omega^1(-;\mathfrak{g})//G$ be the presheaf of $\mathfrak{g}$-valued 1-forms weakly quotiented by the action of the gauge group G. To any presheaf F there is associated a stack, the stackification of F, which is the universal way of making F a stack (Laumon and Moret-Bailly, 2000, Lem. 3.2). $\mathbf{B}G_{\text{conn}}$ is the stackification of $\Omega^1(-;\mathfrak{g})//G$, hence it is a stack. □

In sum, there are separable and nonseparable bundle formulations, and there are separable and nonseparable holonomy formulations. For each bundle formulation there is an equivalent holonomy formulation, and this equivalence respects separability. The association of holonomy representations with nonseparability rests on the coincidental focus in the literature on $\|[\mathbf{P}_1(-), \mathbf{B}G]\|$ as the paradigmatic holonomy representation.

3.4 Conclusion

The primary geographic feature in the interpretive landscape of Yang–Mills theories has been a division between bundle formulations and holonomy formulations. Other theoretical features—determinism, empirical underdetermination, locality, and more—have been mapped along this division. In particular, bundle formulations have generally been treated as separable, while holonomy interpretations have not (Healey, 2007, §2.4). I have argued above that this neglects an important feature of Yang–Mills theories. The distinction between separable and nonseparable theories is unrelated to the distinction between bundles and holonomies. There are separable theories with equivalent configuration stacks $\mathbf{B}G_{\mathrm{conn}}$ and $[\mathbf{P}_1(-), \mathbf{B}G]$, and there are nonseparable theories with equivalent configuration presheaves $\|\mathbf{B}G_{\mathrm{conn}}(-)\|$ and $\|[\mathbf{P}_1(-), \mathbf{B}G]\|$. So even after choosing between bundle and holonomy formulations, we still must choose between separable and nonseparable versions of these.

The foregoing discussion aimed to point out that there are more choices to make than is usually supposed; it did not try to adjudicate this choice. An argument for the untruncated, separable choice would have to appeal to further assumptions about what we are doing when interpreting classical Yang–Mills theory. If we adopt the cost-benefit analysis of the opening paragraph, then the discussion here leads to an argument for the untruncated theory: it is local, and the "surplus" structure is not surplus after all—it represents locality facts (though the precise way in which it represents these locality facts is unclear). If we are motivated by understanding quantized Yang–Mills theory we need some story about the correspondence between the quantum and the classical, along with a sense of whether and how differences in separability make a quantum difference. Benini et al. (2015) argue along these lines, claiming that the separable theory is required if we want to get the global algebra of quantum observables right. Schreiber (2013, §1.1.1), too, has argued that non-perturbative effects in quantum field theory require the structure lost in truncation. Arguments along these lines will be pursued elsewhere.

Finally, note that both separability and the amount of structure turn on the difference between a state space and its truncation. This difference is "internal" to the theory, in the sense that it is a fact about the theory itself, not about how the theory stands in relation to other theories or formulations. To be sure, the groupoid structure of $\mathbf{B}G_{\mathrm{conn}}(X)$ makes a difference to its standing with respect to other theories. It means that $\mathbf{B}G_{\mathrm{conn}}(X)$ has the same amount of structure as $[\mathbf{P}_1(X), \mathbf{B}G]$ and more than $\|[\mathbf{P}_1(X), \mathbf{B}G]\|$, for example. But this groupoid structure also makes $\mathbf{B}G_{\mathrm{conn}}$ separable, and this is just a fact about $\mathbf{B}G_{\mathrm{conn}}$. So if we think that separability is a feature of theories—and it's hard to see what else it could be a feature of—then a theory must be more than its underlying set of models.

Chapter 3, in full, is a reprint of the material as it appears in Dougherty, J. (2017). "Sameness and separability in gauge theories", *Philosophy of Science*, 84(5), 2017. The dissertation author was the sole author of this paper.

4 The hole argument

Developments in the applied sciences give new solutions to some practical problems and streamline old solutions to others. Similarly, developments in the formal sciences give new solutions to some conceptual problems and streamline old solutions to others. The recent confluence of algebraic topology, n-category theory, computer science, logic, and more into a foundational research program called homotopy type theory (HoTT) offers tools for tackling a variety of conceptual problems.[1] Indeed, the subtle notions of sameness at play in homotopy theory and the abstract formality of type theory suggest that these tools can be profitably applied to any problems involving sameness that admit formal statements. In this paper I consider the problem of when two manifolds represent the same spacetime using tools from HoTT.

Specifically, I return to the problems posed by Earman and Norton (1987) in their influential discussion of this problem in the context of modern spacetime theories, particularly GR. They are concerned to show that if you adopt a substantivalist account of spacetime, taking spacetime points to be real, then you are faced with two dilemmas. For if you are a substantivalist, they claim, you must take two Lorentzian manifolds to represent the same spacetime just in case they are identical.[2] But if this is so, any isometric Lorentzian manifold represents an empirically indistinguishable but metaphysically distinct

[1] I will use "HoTT" ecumenically, to refer to any research program interested in a theory the ballpark of the theory presented in the HoTT Book (Univalent Foundations Program, 2013)—which see for an introduction to HoTT. Univalence is the main feature on which my argument depends.

[2] I use the words "identical", "equal", and "the same" interchangeably.

way the world could be. Worse, these worlds are also dynamically indistinguishable, so at any moment there are infinitely many distinct possible futures—a massive failure of determinism. The first problem presses a choice between verificationism and substantivalism and the second a choice between determinism and substantivalism. Even if you are happy to take the substantivalist horn of the verificationist dilemma, the extremity of determinism's failure in the indeterminism dilemma should ward you off from doing the same there.

The most popular defense of substantivalism responds to these dilemmas by supplying a sophisticated modal semantics in which distinct manifolds needn't be distinct spacetimes. I will offer a different account. First, I will appeal to HoTT to argue that even on a straightforward modal semantics the verificationist dilemma does not go through, because it incorrectly distinguishes representations of the same mathematical object. In this I agree with Weatherall (2016b), though the details of our views differ. In particular, I will argue that an argument like Earman and Norton's indeterminism dilemma shows that GR *is* wildly indeterministic when formulated, as is standardly done, using Lorentzian manifolds. But, I claim, a more faithful mathematical representation of GR avoids this problem. This representation employs some n-categorical structure, making HoTT its natural mode of expression.

4.1 Tuples and equality in HoTT

Earman and Norton are concerned with what they call local spacetime theories. A model of such a theory is a tuple of the form $(M, g, g', \dots)$ for some smooth manifold M and geometric structures $g, g', \dots$. The central example in our discussion is GR, which we can formalize by taking its models and laws to be

$$(M, g, T) \qquad \text{and} \qquad G(g) = 8\pi T$$

where g is a Lorentzian metric, T a symmetric $(0, 2)$-tensor, and $G(g)$ the Einstein tensor constructed from g. We can formulate most other field theories using similar kinds of tuples. Much of Earman and Norton's discussion goes through straightforwardly in HoTT, so we will confine our remarks to the most important differences. These are all concerned with equality, which differs in HoTT from the equality relation in first order logic that lies behind Earman and Norton's formalization. Since these differences arise already in simpler non-geometric settings we take one of these (from Weatherall, 2016b) as a toy example. Consider the theory of groups, which we can model with triples $(S, \times, e)$ of a set S, a group operation $\times$ on S, and a unit element e of S. We would like to know when two groups are equal. Two groups $(S, \times, e)$ and $(S', \times', e')$ should be equal if they have the same underlying set equipped with the same group operation and unit element. In this section we make this intuition precise.[3]

Traditionally, mathematics is formalized using first order logic combined with a set theory like Zermelo–Fraenkel set theory. First order logic is untyped, so whether a formula is well-formed depends only on whether every logical operator that appears has been supplied the right number of arguments. HoTT, by contrast, is a type theory, which means that expressions can only be well-formed if they also respect the types of their subexpressions. In this sense, it is less flexible than traditional foundations. For example, consider the following four expressions:

(i) 3 is prime.

(ii) 4 is prime.

(iii) the set of real numbers is prime.

(iv) the set of all sets not containing themselves is prime.

[3] This section is essentially an exposition of equality in Σ-types. For more detail, see the HoTT Book (Univalent Foundations Program, 2013, §1.6).

Traditional foundations and HoTT both agree that (i) is true, (ii) is false, and (iv) is ill-formed. They disagree on (iii): while it is false in traditional foundations—i.e., the set of real numbers is not contained in the extension of the primehood predicate—it is ill-formed in HoTT. This is just one of many calls that a formal system has to make. Any system must count (iv) as ill-formed, on pain of Russell's paradox, but some restrictions are negotiable. While I take it that (iii) is *mathematically* meaningless, it's obviously not a fatal problem if some formal system is liberal enough to grant (iii) a truth value, as the success of traditional foundations shows. HoTT is more restrictive, classing (iii) with (iv). This extra restriction means that the well-formation of some expressions depends on the truth of others: we can only intelligibly ask if x is prime if we know that x is a natural number.

This typing requirement is particularly important for questions of equality. Like every other expression in HoTT, the expression '$x = y$' must be well-typed, and it is only meaningful if x and y have the same type. Similarly, if we would like to ask whether some structure on A is the same as some structure on B, that question depends on A and B being equal. So if we suppose for now that $S = S'$, then we only need to check that the group structures are the same. But this question is only meaningful because we know that $S = S'$, so the answer will depend on this fact. Let $p : S = S'$ mean that the equality of S and S' is witnessed by some p; we then include p in our expressions to mark their dependence on this fact.

The logic of equality-dependence figures into the inference rules of HoTT (Univalent Foundations Program, 2013, §1.12). Equality is reflexive, so for any term 'x' there must be an equality $\mathsf{refl}_x : x = x$ between x and itself. Symmetry and transitivity of equality give new equalities from old. The other rules implement equality's usual substitutional inferential role. Take, for example, the indiscernibility of identicals. For any objects a and a' of type A and predicate P applying to elements of A, if a and a' are equal then

$P(a)$ implies $P(a')$. More generally, suppose that P is instead some construction that takes an object a of type A as input and gives some collection $P(a)$ as output. Then if a and a' are equal, any object in $P(a)$ must be an object in $P(a')$. Using our notation, an equality $p : a = a'$ gives a function $p_* : P(a) \to P(a')$ that sends each object in $P(a)$ to the corresponding object in $P(a')$, obtained by substitution of a' for a according to p. By the rules of equality dependence, the function $(\mathsf{refl}_a)_* : P(a) \to P(a)$ induced by the equality $\mathsf{refl}_a : a = a$ is just the identity function $\mathsf{id}_{P(a)}$.

Applying the indiscernibility of identicals to an equality $p : S = S'$, we can sensibly ask (and answer) questions about whether they are equipped with the same structure. They have the same group operation if $p_*(x \times y) = p_*(x) \times' p_*(y)$ for all x and y in S, and they have the same unit element if $p_*(e) = e'$. In order to complete this approach, then, we need some way of producing an equality of sets. In traditional foundations, this is done with the axiom of extensionality: two sets are the same just in case they have all the same elements. But in HoTT this would require asking whether objects of S are also objects of S', and this question risks being ill-formed. Indeed, since the expression '$x = y$' is only meaningful when x and y have the same type, if x is an element of S and y an element of S' then the meaning of '$x = y$' depends on our already knowing that S and S' are equal, which is our question. And if this expression is meaningless, it can't be used as a consequent in the axiom of extensionality. So phrasing extensionality in this way would be circular.

HoTT answers this question with the univalence axiom, which is an extensionalizing principle of a different kind. Suppose that we had some bijection $f : S \to S'$. Then each element s of S corresponds to a unique element $f(s)$ of S'. And it's hard to say what more we could ask of an equality of type $S = S'$. Any statement about s can be uniquely translated into a statement about $f(s)$, and as such any statement about S can be uniquely translated into one about S' *salva veritate*. This arguably exhausts the content of an equality statement. So, more formally, we suppose[4]

> **The univalence axiom (for sets)**: For every bijection $f : A \to B$ of sets A and B there is an equality $\mathsf{ua}(f) : A = B$ such that $(\mathsf{ua}(f))_* = f$.

So all it takes to get started in showing that $(S, \times, e) = (S', \times', e')$ is to provide a bijection $S \to S'$.

We now have all the tools necessary to show that two groups are equal. Applying this machinery to a concrete example, we can construct an equality of groups $(\mathbb{Z}, +, 0)$ and $(\mathbb{Z}, +', 1)$, where $+'$ is defined by $n +' m = n + m - 1$. This equality is the tuple $(\mathsf{ua}(\phi), \mathsf{refl}_{+'}, \mathsf{refl}_1)$, where $\phi : \mathbb{Z} \to \mathbb{Z}$ is defined by $\phi(n) = n + 1$. This result might surprise you. After all, you might think, 0 and 1 are different integers, and so the tuples $(\mathbb{Z}, +, 0)$ and $(\mathbb{Z}, +', 1)$ differ in their third entries, and so they must be different tuples, hence different groups. But the first of these inferences fails. Recall that to even intelligibly ask the question of whether the third entries of some tuples are the same or different, we must know that they have the same type. And this depends on our knowing that the first entries of each tuple are the same. It depends on it in the strong sense that we *use* the equality of the first entries to compare the third entries. So it's true that $0 \neq 1$, but it's false that $(\mathsf{ua}(\phi))_*(0) \neq 1$, and the second statement is the relevant formalization of the statement "$(\mathbb{Z}, +, 0)$ and $(\mathbb{Z}, +', 1)$ differ in their third entries". Appeal to $0 \neq 1$ shows that there is no equality of the form $(\mathsf{refl}_{\mathbb{Z}}, -)$ between the groups $(\mathbb{Z}, +, 0)$ and $(\mathbb{Z}, +', 1)$. But this is not to say that there is no equality at all.

On a straightforward reading of the HoTT formalism, Earman and Norton's verificationist is similar to the mistaken argument of the previous paragraph; it shows that there is no equality of tuples with a refl in the first entry and concludes from this that the tuples are distinct. This inference fails in the case of GR just as it does in the case of groups. But, perhaps surprisingly, this does not mean that the indeterminism dilemma also

[4] This axiom is apparently weaker than the univalence axiom stated in the HoTT Book (Univalent Foundations Program, 2013, Axiom 2.10.3), but it is equivalent.

fails. The same straightforward reading of HoTT leads right to indeterminism in exactly the places Earman and Norton say it does.

4.2 The verificationist argument

Earman and Norton pitch the verificationist argument as a dilemma for substantivalists: admit observationally indistinguishable but distinct states of affairs or deny substantivalism. However, subsequent philosophical work has shown that this argument is really better conceived of as a *reductio*, for there is more than one way to avoid both horns of the dilemma. Nevertheless, some ways have gone mostly unexplored. This is a shame if, like me, your intuitive response is one of these. In this section, I use HoTT to articulate one intuitive response to the verificationist argument and defend this response against a *prima facie* objection to its coherence. I take the initial plausibility of this objection to explain the unpopularity of responses like mine in the philosophical literature.

Here is one way to state the verificationist argument. Suppose we take a model of GR to be a pair (M, g) of a smooth manifold M and a Lorentzian metric g on M, and let $\phi : M \to M$ be any nontrivial diffeomorphism. Then we have the following *reductio*:

(1) Substantivalism is true.

(2) By (1), distinct mathematical objects represent distinct possible states of affairs.

(3) (M, g) and $(M, \phi_* g)$ are distinct mathematical objects.

(4) By (2) and (3), (M, g) and $(M, \phi_* g)$ represent distinct possible states of affairs.

We have two reasons to reject this conclusion. First, everyone agrees that (M, g) and $(M, \phi_* g)$ are observationally indistinguishable, since this is what the diffeomorphism ϕ represents. But if (4) is right, then substantivalism leads to distinct but observationally indistinguishable states of affairs. And this raises epistemic concerns about what reasons could underwrite our belief that these unobservable distinctions exists. But substantivalists have lived with this problem since the beginning of (Newtonian) time, and I don't suppose that this argument poses any new challenge. A second, more worrisome problem is that taking (M, g) and $(M, \phi_* g)$ to be inequivalent situations is "at odds with standard modern texts in general relativity, in which this equivalence is accepted unquestioningly in the specific case of manifolds with metrics" (Earman and Norton, 1987, p. 522). So while substantivalists might be able to avoid the first worry, this argument suggests that we have to choose between substantivalism and accord with the textbooks. This distinction is also the seed of the indeterminism argument of the next section, which is more devastating. So we have to reject one of (1)–(3), and since (1) looks the most controversial, it's on the chopping block.

However, we might reject steps other than (1). Substantivalism is a metaphysical thesis about spacetime, and the conclusion of (2) is a thesis about how our mathematical representations relate to modal distinctness facts. This inference must be underwritten by some story about how mathematical representations relate to modal distinctness facts, and how these distinctness facts relate to a mere existence claim about spacetime. Perhaps on a straightforward reading (2) follows from (1), but it needn't. There are various versions of substantivalism that are realist about spacetime points while regarding (M, g) and $(M, \phi_* g)$ as two representations of the same possible world; call these "sophisticated", following Belot and Earman (1999, 2001). The details of such a view can be worked out in a number of ways (e.g., Butterfield, 1989; Brighouse, 1994; Hoefer, 1996; Pooley, 2006), but sophisticated substantivalists are united in their denial that distinctions in the mathematics must be reflected in the existence of corresponding metaphysically distinct states of affairs.

I do not disagree that (2) can be rejected, but developing a more sophisticated account of substantivalism seems to me an overly complicated response to the verificationist argument. Intuitively, (3) is false, and there is no nearby true premise that can rescue the argument. And, as Shulman (2015) has also argued, this intuition is borne out in HoTT. By univalence there is an equality $\mathsf{ua}(\phi) : M = M$, and we have $(\mathsf{ua}(\phi))_*(g) = \phi_* g$ by the defining property of $\mathsf{ua}(\phi)$ and the way that diffeomorphisms act on Lorentzian metrics (i.e., by pushforward). So we have an equality $(\mathsf{ua}(\phi), \mathsf{refl}_{\phi_* g})$ between (M, g) and $(M, \phi_* g)$, and if they are equal they cannot be distinct, so (3) fails. HoTT allows us to keep a straightforward modal semantics without falling prey to the verificationist argument.

This response to the verificationist argument is as short and simple as it is intuitive. Granted, this intuition may depend on upbringing. But the contrast between its simplicity and unpopularity is somewhat puzzling. The culprit is probably an argument similar to the mistaken argument about groups in the previous section: $g \neq \phi_* g$, so $(M, g) \neq (M, \phi_* g)$. But this argument fails for the same reason it did for groups. What *really* follows from the first fact is that there is no equality $(M, g) = (M, \phi_* g)$ of the form $(\mathsf{refl}_M, -)$. But this is not to say that there is no equality at all, and indeed there is one, as we saw in the last paragraph.

There is, it's true, something compelling about this argument from the fact that $g \neq \phi_* g$. And it would be unfair to saddle my opponent with a formalism that simply defines away their argument. But it is possible to reconstruct a version of the verificationist argument in HoTT that underwrites the defense of the previous paragraph. Indeed, it is illuminating to do so, for it shows just where informal versions of it go astray. Suppose that there is some set P containing every spacetime point in every possible world (or names for such), and suppose that a model of GR is given by a triple (M, g, n) of a smooth manifold M, a Lorentzian metric g, and an injection $n : M \to P$ that takes each point p of M to itself (or its name) in P. Then an equality between (M, g, n) and $(M, \phi_* g, n)$ is a diffeomorphism $\psi : M \to M$ such that $\psi_* g = \phi_* g$ and $n \circ \psi = n$. Since n is an injection, the second condition implies that ψ is the identity map. So if we model the spacetimes of GR with triples like this, then the only equality $M = M$ that could underlie an equality of our models is refl_M. Thus if we replace (M, g) and $(M, \phi_* g)$ with (M, g, n) and $(M, \phi_* g, n)$ in step (3) in the verificationist argument, then the modified premise is true. But these "labeled Lorentzian manifolds" aren't the models of GR, so there's no particular reason to worry about the conclusion of the argument on this reading. The argument only goes through in a worrisome way if we switch formalizations halfway through, using triples like (M, g, n) in step (3) but pairs like (M, g) in step (4). This mistake might be particularly tempting in traditional foundations, where (M, g) and $(M, \phi_* g)$ are merely isomorphic as Lorentzian manifolds and not equal as ordered tuples, but this doesn't mean we must make such mistakes illegal by adopting HoTT, we only need to be more careful.

I take Weatherall (2016b) to be making this foundation-agnostic argument. He explicitly adopts two positions on the application of mathematics to physics: that "the default sense of 'sameness' or 'equivalence' of mathematical models in physics should be the sense of equivalence given by the mathematics" and that "the standard of sameness for mathematical objects is some form of isomorphism" (Weatherall, 2016b, p. 3). Since (M, g) and $(M, \phi_* g)$ are isomorphic as Lorentzian manifolds, these positions imply that (M, g) and $(M, \phi_* g)$ should be the same for the purposes of mathematical physics. And so, like me, Weatherall denies (3) in the verificationist argument. Indeed, something like his first principle is implicit in my taking equal Lorentzian manifolds in HoTT to represent the same spacetimes, and univalence is something like his second principle.

But I do want to claim something stronger than Weatherall. HoTT might not be necessary to deny (3), but adopting a straightforward reading of HoTT involves more than just the rejection of (3). Suppose, for example, that we accepted Weatherall's principles but opted for traditional foundations. We might adopt a form of sophisticated substantivalism like Pooley's (Pooley, 2006, p. 103) on which "individual spacetime points exist as basic objects, but possible spacetimes correspond to equivalence classes of diffeomorphic models of GR". Then isomorphic models like (M, g) and $(M, \phi_* g)$ correspond to the same spacetime, in accordance with our principles, even when they are unequal. But precisely because they are unequal, it takes some work to argue that this distinction is irrelevant for our purposes. That is, the opponent of this response will demand an account of why the term “distinct” in (3) should not be read strictly as “unequal”. It seems difficult to give such an account without adverting to the kind of sophisticated semantics used to reject (2). And even if we succeed, we are left with the demand from the philosopher of mathematics for an account of why the standard of sameness for mathematical objects isn’t equality and, furthermore, of what role equality plays if that is the case (Pooley, 2006, p. 103). HoTT doesn’t have these problems.

This is not to say that a straightforward HoTT story like mine doesn’t face questions; it just faces different ones. So we can agree on Weatherall’s principles while disagreeing on other issues and facing different problems. In the next two sections I want to focus on a disagreement between my HoTT-based account and those like Pooley’s which take possible spacetimes to correspond to equivalence classes of diffeomorphic models of GR. One virtue of sophisticated accounts is that they avoid a second problem that threatens to sink substantivalism: wild and intuitively spurious indeterminism. On my HoTT-based picture, GR formulated using Lorentzian manifolds *is* indeterministic in this way. But I take this to be a virtue of the picture. There is a natural way to rectify this indeterminism, and if we do so then I claim we more faithfully capture textbook GR.

4.3 The indeterminism argument

The real star of Earman and Norton’s show is the indeterminism dilemma, which transposes the theme of underdetermination due to diffeomorphism from an epistemic to a metaphysical key. Sophisticated substantivalism has a blanket response to both problems: distinctions in the mathematics don’t force distinctions in the metaphysics. If we use a straightforward semantics for HoTT, though, this blanket response isn't available. Indeed, even though the verificationist argument doesn't go through on a HoTT reading, the indeterminism argument does.

Determinism in our context is a claim about models satisfying the physical laws. As a first pass, a theory should be deterministic if any dynamically possible initial datum has a unique dynamically possible extension.[5] If GR is deterministic, we should expect that fixing a dynamically possible state of the world everywhere outside of some "nice" region U should also suffice to fix a unique dynamically possible state of the world inside of U. Tolerable levels of indeterminism vary from person to person, thus so will the bounds of "niceness". But surely a unit ball centered on the origin in Minkowski spacetime is about as nice as can be, and if some dynamically possible universe is Minkowski outside this ball it ought to be Minkowski inside, as well. So consider the exterior specification $(\mathbb{R}^4, \eta|_{B^c})$, where η is the Minkowski metric on $\mathbb{R}^4$ and B^c is $\mathbb{R}^4$ without the unit ball centered on the origin. GR is deterministic only if there is a unique extension of this initial datum. Since Minkowski space $(\mathbb{R}^4, \eta)$ is such an extension, we really only need to be concerned with uniqueness of the extension, not existence.

Earman and Norton's indeterminism argument concludes that uniqueness is undermined for substantivalists by "hole diffeomorphisms", which differ from the identity only on some nice region. In our example, if $\phi : \mathbb{R}^4 \to \mathbb{R}^4$ is some diffeomorphism supported only in the unit ball then $\phi_* \eta|_{B^c} = \eta|_{B^c}$. Thus $(\mathbb{R}^4, \eta)$ and $(\mathbb{R}^4, \phi_* \eta)$ are different extensions of the initial datum $(\mathbb{R}^4, \eta|_{B^c})$ to the entire spacetime. And since the dynamics of GR are diffeomorphism-invariant, they are both dynamically possible extensions. So there is no unique way to extend this datum. Generalizing this example, Earman and Norton claim

[5] This is a first pass for a couple reasons. For one, the notion of "initial datum" is vague, though in this specific instance I expect all will agree that a full specification of the configuration outside a small hole will qualify as enough initial data. The more vexatious problem is a suite of counterexamples to the "possibility counting" notion of determinism that I am using (Belot, 1995; Brighouse, 1997; Melia, 1999). I argue elsewhere that the stronger sense of uniqueness we are about to encounter can deal with these cases.

that "the state within *any* neighborhood of the manifold can *never* be determined by the state exterior to it, no matter how small the neighborhood and how extensive the exterior specification" (Earman and Norton, 1987, p. 524). Indeterminism this radical suggests that something has gone wrong in our understanding, given GR's success in predicting gravitational phenomena.

As we saw in the last section, a sophisticated substantivalist like Pooley has an easy reply to this: possible spacetimes are given by isomorphism classes of Lorentzian manifolds, so $(\mathbb{R}^4, \eta)$ and $(\mathbb{R}^4, \phi_*\eta)$ aren't different spacetimes, so Earman and Norton's argument fails. It might seem as though the straightforward HoTT account can reply to Earman and Norton's argument in just the same way. And indeed, as just stated the argument fails, because $(\mathbb{R}^4, \eta)$ and $(\mathbb{R}^4, \phi_*\eta)$ are not distinct.

But this doesn't completely save the straightforward HoTT account from the counterexample. Recall that, in HoTT, reasoning with equalities requires an appeal to particular equalities. The indiscernibility of identicals is an inference rule that takes four things as inputs: objects a and a' of type A, a predicate P that applies to objects of type A, and an equality $p : a = a'$. In order to conclude $p_* : P(a) \to P(a')$, we must have a particular p on which to rely. Similarly, sentences involving equality statements refer to particular equalities. So when we demand that some initial datum have a unique dynamically possible extension—which is to say that an extension exists and any two extensions are equal—we mean that an extension exists and there is a *particular* equality between any two extensions. And since equal initial data must have equal dynamically possible extensions, showing that GR is deterministic requires showing that for any particular equality of initial data there is a particular extension of that equality to an equality of dynamically possible extensions.

The example above shows that there is an equality of initial data that cannot be extended to an equality of dynamically possible extensions. For consider equalities of type $(\mathbb{R}^4, \eta|_{B^c}) = (\mathbb{R}^4, \phi_*\eta|_{B^c})$. These are diffeomorphisms $\psi : \mathbb{R}^4 \to \mathbb{R}^4$ such that

$\psi_*\eta|_{B^c} = \phi_*\eta|_{B^c}$. So one equality of this type is $(\mathsf{refl}_{\mathbb{R}^4}, \mathsf{refl}_{\eta|_{B^c}})$. But this equality does not generally extend to an equality between $(\mathbb{R}^4, \eta)$ and $(\mathbb{R}^4, \phi_*\eta)$. In general, the only equality between these manifolds is $(\mathsf{ua}(\phi), \mathsf{refl}_{\phi_*\eta})$, and this is not an extension of $(\mathsf{refl}_{\mathbb{R}^4}, \mathsf{refl}_{\eta|_{B^c}})$ because they have different underlying diffeomorphisms. So on the natural reading of uniqueness in HoTT, there is not a unique extension of any initial datum to a complete spacetime.[6]

There are two ways to respond to this problem: we can modify our definition of determinism, or we can modify what we take models of GR to be. I think the former isn't promising, for two reasons. First, there is good reason to think that the stronger requirement is generally the right way to formalize uniqueness statements in HoTT. In abstract homotopy theory and n-category theory, which are two ingredients of the HoTT program, the correct notion of uniqueness corresponds to the stronger requirement. Second, many of the difficult cases for defining determinism in light of the hole argument (found in, e.g., Melia, 1999) pose problems for the "mere existence" definition that are avoided by the "particular equality" definition. So let's consider the second route. The problem with Lorentzian manifolds, according to the argument just given, is that there aren't quite enough equalities for uniqueness to hold. The counterexample spacetimes aren't *distinct*, there just aren't enough equalities between them to support all the extensions we need. The quick fix is to simply declare that all equalities of initial data to extend to equalities of spacetimes. I take this to produce a better formalization of textbook GR.

[6] In HoTT vocabulary, this shows that while the type of dynamically possible spacetimes extending some dynamically possible initial datum is *connected* (Univalent Foundations Program, 2013, Def. 7.5.1), it is not *contractible* (Univalent Foundations Program, 2013, Def. 3.11.1).

4.4 Generally covariant Lorentzian manifolds

We can regain determinism in GR by inserting the equalities that seem to be missing. This is a kind of quotient, since we are identifying possible futures if they restrict to equal initial data. This might sound like the version of sophisticated substantivalism that takes spacetimes to be given by diffeomorphism classes of Lorentzian manifolds. But it is importantly different. Taking equivalence classes in the usual sense eliminates any nontrivial equality dependence, since there is at most one equality between any two equivalence classes. If we apply a weaker quotient by inserting more equalities, then we can avoid the indeterminism of Lorentzian manifolds without completely collapsing the equality structure. This equality structure is used in GR, so we ought to use these weakly quotiented models, instead.

A precise statement of the quotient procedure would require us to identify exactly which diffeomorphisms cause problems for determinism. I will call these hole diffeomorphisms, and I will take them to at least include those diffeomorphisms generated by a vector field. Exactly which diffeomorphisms are hole diffeomorphisms is a substantive question beyond the scope of this article. The result of the quotient is the following

Definition 4.4.1. A *generally covariant Lorentzian metric* on a manifold M is specified by a Lorentzian metric g, and an equality $g = g'$ of generally covariant Lorentzian metrics on M is a hole diffeomorphism $\psi : M \to M$ such that $\psi_* g = g'$. A *generally covariant Lorentzian manifold* is a pair (M, g) of a smooth manifold M and a generally covariant Lorentzian metric g on M. Thus, an equality $(M, g) = (M', g')$ of generally covariant Lorentzian manifolds is a pair (ϕ, ψ) of a diffeomorphism $\phi : M \to M'$ and a "hole" diffeomorphism $\psi : M' \to M'$ such that $\psi_* \phi_* g = g'$.

A generally covariant Lorentzian manifold is not the same thing as a Lorentzian manifold. For example, for any hole diffeomorphism $\phi : M \to M$ there is an equality between (M, g)

and $(M, \phi_* g)$ given by (refl_M, ϕ). So there *is* an equality of the form $(\mathsf{refl}_M, -)$, which there is not for Lorentzian manifolds, as we saw in considering the verificationist argument. And this allows generally covariant Lorentzian manifolds to avoid the refined indeterminism argument of the previous section. The problem, recall, was that there was no way to extend the equality $(\mathsf{refl}_{\mathbb{R}^4}, \mathsf{refl}_{\eta|_{B^c}})$ to an equality between $(\mathbb{R}^4, \eta)$ and $(\mathbb{R}^4, \phi_* \eta)$ that had $\mathsf{refl}_{\mathbb{R}^4}$ as its underlying equality of manifolds. But now there is: we can just take $(\mathsf{refl}_{\mathbb{R}^4}, \phi)$. And indeed, one can show that any equality of initial data extends to an equality of generally covariant Lorentzian manifolds, so determinism is regained and the verificationist argument still fails at the third step.[7]

I take generally covariant Lorentzian manifolds to be a better formalization of models of GR than Lorentzian manifolds. This isn't only because they avoid the indeterminism of the refined indeterminism argument (though that helps), but also because they more accurately capture GR as found in textbooks. The difference between the two can be hard to find. While the indeterminism argument distinguishes them, the difference doesn't much arise in everyday GR life. And for good reason: generally covariant Lorentzian manifolds are obtained by quotienting the type of Lorentzian manifolds, and we generally work with quotients by doing manipulations on the unquotiented objects and then showing that these manipulations are preserved by the quotient. We think nothing of treating a fraction as a pair of integers, even though it's really an equivalence class of such. We just make sure not to manipulate them in a way that violates the equivalence relation: we allow definitions like $a/b + c/d = (ad + bc)/bd$ but not like $a/b \star c/d = (a + c)/(b + d)$. Since any Lorentzian manifold gives a generally covariant Lorentzian manifold, and any equality of Lorentzian manifolds gives an equality of generally covariant Lorentzian manifolds, we can usually work with regular Lorentzian manifolds, just as we work with pairs of integers when treating fractions. Moreover, if we have an equality $(\mathsf{ua}(\phi), \psi) : (M, g) =$

[7] The claim that such an extension is possible is essentially the claim that the initial value problem of GR is well-posed (Wald, 1984, Th. 10.2.2).

(M', g') of generally covariant Lorentzian manifolds then there is a corresponding equality $(\mathsf{ua}(\psi \circ \phi), \mathsf{refl}_{g'}) : (M, g) = (M', g')$ of Lorentzian manifolds, so if we only care about whether there is some equality or other between two objects, Lorentzian manifolds and generally covariant Lorentzian manifolds amount to the same thing. We can only see the difference if we pay attention to particular equalities.[8]

Though the differences are inconspicuous in many contexts, they are crucial in others. The defining difference is that for a fixed manifold, two generally covariant metrics g and g' can be equal in more than one way. This feature is most familiar from classical gauge theories, and so we should expect the difference between Lorentzian manifolds and their generally covariant cousins to manifest wherever gauge structure manifests—e.g., in perturbation theory, in initial value problems, and in the structure of fields "at infinity". Indeed, the claim that "diffeomorphisms comprise the gauge freedom of general relativity" (Wald, 1984, p. 438) is textbook GR, and this freedom is found in all three of these situations (Wald, 1984, Eq. 4.4.9, Eq. 10.2.32, p. 467 fn. 2). Consider the first.

In the weak-field approximation of GR, we consider the infinitesimal neighborhood of a flat model like Minkowski space $(\mathbb{R}^4, \eta)$. To do this, we fix a background manifold like $\mathbb{R}^4$ and some metric like η, and we parametrize the space of configurations on $\mathbb{R}^4$ by writing any metric g as $g = \eta + \gamma$. If we are working with Lorentzian metrics, then the equality facts in this space are trivial: two metrics $\eta + \gamma$ and $\eta + \gamma'$ are equal just in case $\gamma = \gamma'$. On the other hand, if we are working with generally covariant Lorentzian metrics, then the equality facts in this space are more complicated: an equality of metrics $\eta + \gamma$ and $\eta + \gamma'$ is a hole diffeomorphism $\phi : \mathbb{R}^4 \to \mathbb{R}^4$ such that $\phi_*(\eta + \gamma) = \eta + \gamma'$. This parametrizes the neighborhood of Minkowski space, so to restrict attention to infinitesimal perturbations we demand that γ be infinitesimal. If we are working with Lorentzian metrics, then the

[8] In the categorical semantics, the groupoid of Lorentzian manifolds and the groupoid of generally covariant Lorentzian manifolds have the same objects, and the identity functor induces a bijection on isomorphism classes. However, it does not induce a bijection on hom-sets, so the categories are not equivalent.

equality facts don't change: two metrics are equal just in case $\gamma = \gamma'$. But if we are working with generally covariant Lorentzian metrics, then the equality facts must also become infinitesimal. So an equality between $\eta + \gamma$ and $\eta + \gamma'$ with γ and γ' infinitesimal is given by a vector field ξ such that $\gamma + \pounds_\xi \eta = \gamma'$, where $\pounds_\xi$ is the Lie derivative along ξ. To decide which of these is the correct space of infinitesimal perturbations off of Minkowski space we just need to check what the equality facts about perturbations are according to a GR textbook. As Wald (Wald, 1984, Eq. 4.4.9) says, a gauge transformation (in our parlance, an equality) in this context is a transformation taking γ to $\gamma + \pounds_\xi \eta$. So the correct space of perturbations comes from generally covariant Lorentzian metrics.[9]

4.5 Conclusion

I have advanced two claims. First, HoTT gives us the technology to articulate an intuitive response to the verificationist argument that does not provoke the same *prima facie* difficulties that arise when this response is stated using traditional foundations of mathematics. This is not intended to be a technical solution to a philosophical problem, but a way to streamline an available philosophical solution. I take the primary obstacle to rejecting the distinctness claim (3) of the verificationist argument to be merely technical: as an artifact of traditional foundations, we have to use multiple representations of a single mathematical object. We can reject (3) without appealing to HoTT, and it's been done (Weatherall, 2016b). So when it comes to the verificationist argument, HoTT is a way to avoid a technical limitation.

The second claim is meant to be more substantive. Responding to the technical limitation of the verificationist argument with HoTT motivates further philosophical

[9] This discussion of perturbation theory in GR in HoTT is admittedly far too brief to be convincing. I gesture at it only to indicate that general covariance in my sense has substantive consequences and they are the correct ones. A detailed treatment of perturbation theory, as well as initial value problems and boundary conditions, will be given elsewhere.

conclusions. In particular, if we take models of GR to be Lorentzian manifolds then on a natural reading the HoTT-motivated account takes GR to be badly indeterministic on account of hole diffeomorphisms. The natural solution is to move to what I have called generally covariant Lorentzian manifolds. On this reconstruction, the hole argument shows that any diffeomorphism-invariant dynamics on a regular Lorentzian manifold will lead to indeterminism, and that a substantive notion of general covariance rectifies this problem. I take this to be an advantage of my account insofar as it corroborates the feeling many physicists have that general covariance is a substantive notion, and that the hole argument tells us something important about it (Belot and Earman, 1999). But my primary argument for taking generally covariant Lorentzian manifolds to be the correct model of GR is appeal to the physics. In most cases, the difference between generally covariant Lorentzian manifolds and their plainer cousins is irrelevant. But when the difference makes a difference—for example, in perturbation theory—physicists use the generally covariant structure.

Chapter 4, in full, is a reprint of material currently being prepared for submission for publication. The dissertation author is the sole author of this material.

Bibliography

Adams, R. M. (1979). Primitive thisness and primitive identity. *The Journal of Philosophy*, 76(1):5–26.

Aharonov, Y. and Bohm, D. (1959). Significance of electromagnetic potentials in the quantum theory. *Physical Review*, 115:485–491.

Anderson, E. (1995). Feminist epistemology: An interpretation and a defense. *Hypatia*, 10(3):50–84.

Baez, J. and Muniain, J. P. (1994). *Gauge Fields, Knots and Gravity*. World Scientific.

Baez, J. C. (2009). Torsors made easy. http://math.ucr.edu/home/baez/torsors.html.

Baez, J. C. and Shulman, M. (2010). Lectures on n-categories and cohomology. In Baez, J. C. and May, J. P., editors, *Towards Higher Categories*, pages 1–68. Springer New York.

Baker, D. J. (2010). Symmetry and the metaphysics of physics. *Philosophy Compass*, 5(12):1157–1166.

Balachandran, A. P. (1994). Gauge symmetries, topology, and quantization. *AIP Confrence Proceedings*, 317:1–81.

Barrett, T. W. (2018). What do symmetries tell us about structure? *Philosophy of Science*.

Barwick, C. (2017). The future of homotopy theory. http://www.maths.ed.ac.uk/~cbarwick/papers/future.pdf.

Belot, G. (1995). New work for counterpart theorists: Determinism. *The British Journal for the Philosophy of Science*, 46(2):185–195.

Belot, G. (1998). Understanding electromagnetism. *The British Journal for the Philosophy of Science*, 49(4):531–555.

Belot, G. (2001). The principle of sufficient reason. *The Journal of Philosophy*, 98(2):55–74.

Belot, G. (2003). Notes on symmetries. In Brading and Castellani (2003b), pages 393–412.

Belot, G. (2007). The representation of time and change in mechanics. In Earman, J. and Butterfield, J., editors, *Philosophy of Physics*, pages 133–227. North-Holland.

Belot, G. (2013). Symmetry and equivalence. In Batterman, R., editor, *The Oxford Handbook of Philosophy of Physics*, pages 318–339. Oxford University Press.

Belot, G. (2018). Fifty million Elvis fans can't be wrong. *Noûs*.

Belot, G. and Earman, J. (1999). From metaphysics to physics. In Butterfield, J. and Pagonis, C., editors, *From Physics to Philosophy*, chapter 7, pages 166–186. Cambridge University Press.

Belot, G. and Earman, J. (2001). Pre-Socratic quantum gravity. In *Physics Meets Philosophy at the Planck Scale: Contemporary Theories in Quantum Gravity*, chapter 10, pages 213–255. Cambridge University Press.

Benini, M., Schenkel, A., and Szabo, R. (2015). Homotopy colimits and global observables in abelian gauge theory. *Letters in Mathematical Physics*, 105(9):1193–1222.

Black, M. (1952). The identity of indiscernibles. *Mind*, 61(242):153–164.

Brading, K. and Castellani, E. (2003a). Introduction. In Brading and Castellani (2003b), pages 1–18.

Brading, K. and Castellani, E., editors (2003b). *Symmetries in Physics: Philosophical Reflections*. Cambridge University Press.

Brighouse, C. (1994). Spacetime and holes. In *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, volume 1, pages 117–125. Philosophy of Science Association.

Brighouse, C. (1997). Determinism and modality. *The British Journal for the Philosophy of Science*, 48(4):465.

Brown, R. (1987). From groups to groupoids: a brief survey. *Bulletin of the London Mathematical Society*, 19(2):113–134.

Butler, J. (1736). *Fifteen Sermons Preached at the Rolls Chapel*. Hilliard and Brown.

Butterfield, J. (1989). The hole truth. *The British Journal for the Philosophy of Science*, 40(1):1–28.

Butterfield, J. (2006). On symplectic reduction in classical mechanics. In Butterfield, J. and Earman, J., editors, *The Handbook of Philosophy of Physics*, pages 1–131. North Holland.

Callan, C. G., Dashen, R. F., and Gross, D. J. (1976). The structure of the gauge theory vacuum. *Physics Letters*, 63B(3):334–340.

Carnap, R. (1956). The methodological character of theoretical concepts. In Feigl, H. and Scriven, M., editors, *Foundations of Science & the Concepts of Psychology and Psychoanalysis*, volume I of *Minnesota Studies in Philosophy of Science*, pages 38–76. University of Minnesota Press.

Carnap, R. (1966). *Philosophical Foundations of Physics: An Introduction to the Philosophy of Science*. Basic Books.

Caulton, A. (2013). Discerning "indistinguishable" quantum systems. *Philosophy of Science*, 80:49–72.

Caulton, A. (2015). The role of symmetry in the interpretation of physical theories. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 52:153–162.

Chakravartty, A. (2007). *A Metaphysics for Scientific Realism*. Cambridge University Press.

Chang, H. (2004). *Inventing Temperature: Measurement and Scientific Progress*. Oxford University Press.

Coleman, S. (1985). The uses of instantons. In *Aspects of Symmetry*, chapter 7, pages 265–350. Cambridge University Press.

Dasgupta, S. (2016). Symmetry as an epistemic notion (twice over). *The British Journal for the Philosophy of Science*, 67:837–878.

Dasgupta, S. (2018). Symmetry and superfluous structure: A metaphysical overview. To appear in Knox, E. and Wilson, A., editors, *The Routledge Companion to Philosophy of Physics*.

Dewar, N. (2017). Sophistication about symmetries. *The British Journal for the Philosophy of Science*. axx021.

Dewar, N. (2018). Maxwell gravitation. *Philosophy of Science*, 85(2):249–270.

Duhem, P. (1914). *La Théorie Physique. Son Objet, sa Structure*. Marcel Rivière et Cie. Published in English translation as Duhem (1962).

Duhem, P. (1962). *The Aim and Structure of Physical Theory*. Atheneum.

Dwyer, W. G., Hirschhorn, P. S., Kan, D. M., and Smith, J. H. (2004). *Homotopy Limit Functors on Model Categories and Homotopical Categories*. American Mathematical Society.

Earman, J. (1989). *World Enough and Space-Time*. Cambridge University Press.

Earman, J. and Norton, J. D. (1987). What price spacetime substantivalism? The hole story. *The British Journal for the Philosophy of Science*, 38(4):515–525.

Ehrenfest-Afanassjewa, T. (1925). Zur Axiomatisierung des zweiten Hauptsatzes der Thermodynamik. *Zeitschrift für Physik*, 33:933–945.

Enderton, H. B. (1977). *Elements of Set Theory*. Academic Press.

Fort, H. and Gambini, R. (2000). U(1) puzzle and the strong CP problem from a holonomy perspective. *International Journal of Theoretical Physics*, 39:341–349.

Frege, G. (1892). Über Sinn und Bedeutung. *Zeitschrift für Philosophie und philosophische Kritik*, 100:25–50. Published in English translation as Frege (1948).

Frege, G. (1948). Sense and reference. *The Philosophical Review*, 57(3):209–230.

French, S. (1995). Hacking away at the identity of indiscernibles: Possible worlds and Einstein's principle of equivalence. *The Journal of Philosophy*, 92(9):455–466.

French, S. (2014). *The Structure of the World*. Oxford University Press.

French, S. and Redhead, M. (1988). Quantum physics and the identity of indiscernibles. *British Journal for the Philosophy of Science*, 39:233–246.

Galilei, G. (1967). *Dialogue Concerning the Two Chief World Systems—Ptolemaic & Copernican*. University of California Press. Trans. S. Drake.

Giraud, J. (1971). *Cohomologie non abélienne*, volume 179 of *Die Grundlehren der mathematischen Wissenschaften*. Springer.

Giulini, D. (1995). Asymptotic symmetry groups of long-ranged gauge configurations. *Modern Physics Letters*, 10:2059–2070.

Goerss, P. G. and Jardine, J. F. (1999). *Simplicial Homotopy Theory*, volume 147 of *Progress in Mathematics*. Birkhäuser.

Greaves, H. and Wallace, D. (2014). Empirical consequences of symmetries. *The British Journal for the Philosophy of Science*, 65(1):59–89.

Gribov, V. N. (1978). Quantization of non-Abelian gauge theories. *Nuclear Physics B*, 139(1):1–19.

Hacking, I. (1975). The identity of indiscernibles. *The Journal of Philosophy*, 72(9):249–256.

Halvorson, H. (2012). What scientific theories could not be. *Philosophy of Science*, 79(2):183–206.

Hawley, K. (2009). Identity and indiscernibility. *Mind*, 118(469):101–119.

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

Healey, R. (2009). Perfect symmetries. *The British Journal for the Philosophy of Science*, 60(4):697–720.

Hirschhorn, P. S. (2003). *Model Categories and Their Localizations*. American Mathematical Society.

Hoefer, C. (1996). The metaphysics of space-time substantivalism. *The Journal of Philosophy*, 93(1):5–27.

Hollander, S. (2008). A homotopy theory for stacks. *Israel Journal of Mathematics*, 163(1):93–124.

Hubbard, J. H. and Hubbard, B. B. (2001). *Vector Calculus, Linear Algebra, and Differential Forms: A Unified Approach*. Prentice Hall, second edition.

Huggett, N. (1999). On the significance of permutation symmetry. *British Journal for the Philosophy of Science*, 50:325–347.

Ismael, J. and van Fraassen, B. C. (2003). Symmetry as a guide to superfluous theoretical structure. In Brading and Castellani (2003b), pages 371–392.

Jackiw, R. (1980). Introduction to the Yang–Mills quantum theory. *Reviews of Modern Physics*, 52(4):661–673.

Jackiw, R. and Rebbi, C. (1976). Vacuum periodicity in the Yang–Mills quantum theory. *Physical Review Letters*, 37(3):172–175.

Jackson, J. D. (1999). *Classical Electrodynamics*. John Wiley & Sons.

Knox, E. (2014). Newtonian spacetime structure in light of the equivalence principle. *The British Journal for the Philosophy of Science*, 65(4):863–880.

Kolář, I., Michor, P. W., and Slovák, J. (1993). *Natural Operations in Differential Geometry*. Springer-Verlag.

Kosso, P. (2000). The empirical status of symmetries in physics. *The British Journal for the Philosophy of Science*, 51(1):81.

Kripke, S. (1980). *Naming and Necessity*. Harvard University Press.

Ladyman, J. and Presnell, S. (2017). Identity in homotopy type theory: Part ii, the conceptual and philosophical status of identity in hott†. *Philosophia Mathematica*, 25(2):210–245.

Landau, L. D. and Lifshitz, E. M. (1980). *The Classical Theory of Fields*. Butterworth–Heinemann, fourth revised English edition. Trans. M. Hamermesh.

Landini, G. and Foster, T. R. (1991). The persistence of counterexample: Re-examining the debate over Leibniz law. *Noûs*, 25(1):43–61.

Lange, M. (2002). *An Introduction to the Philosophy of Physics: Locality, Fields, Energy, and Mass*. Blackwell.

Lárusson, F. (2009). The homotopy theory of equivalence relations. arXiv:math/0611344.

Laudan, L. (1990). Demystifying underdetermination. In Savage, C. W., editor, *Scientific Theories*, volume XIX of *Minnesota Studies in Philosophy of Science*, pages 267–297. University of Minnesota Press.

Laumon, G. and Moret-Bailly, L. (2000). *Champs algébriques*. Springer.

Leibniz, G. W. and Clarke, S. (2000). *Correspondence*. Hackett.

Lewis, D. (1970). How to define theoretical terms. *The Journal of Philosophy*, 67(13):427–446.

Lewis, D. (1986). *On the Plurality of Worlds*. Blackwell.

Lewis, D. (2009). Ramseyan humility. In Braddon-Mitchell, D. and Nola, R., editors, *Conceptual Analysis and Philosophical Naturalism*, pages 203–222. MIT Press.

Longino, H. (1990). *Science as Social Knowledge*. Princeton University Press.

Lurie, J. (2009). *Higher Topos Theory*. Princeton University Press.

Lyre, H. (1999). Gauges, holes, and their 'connections'. arXiv:gr-qc/9904036.

Lyre, H. (2004). Holism and structuralism in $U(1)$ gauge theory. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 35(4):643–670.

Mac Lane, S. and Moerdijk, I. (1992). *Sheaves in Geometry and Logic*. Springer.

Malament, D. B. (2012). *Topics in the Foundations of General Relativity and Newtonian Gravitation Theory*. University of Chicago Press.

Malaney, P. (1996). *The Index Number Problem: A Differential Geometric Approach*. PhD thesis, Harvard University.

Maldacena, J. (2015). The symmetry and simplicity of the laws of physics and the Higgs boson. *European Journal of Physics*, 37(1):015802.

Martin-Löf, P. (1975). An intuitionistic theory of types: predicative part. In Rose, H. E. and Shepherdon, J. C., editors, *Logic Colloquium*, pages 73–118. North Holland.

Maudlin, T. (2002). Thoroughly muddled McTaggart: Or, how to abuse gauge freedom to create metaphysical monostrosities. *Philosophers' Imprint*, 2(4):1–23.

May, J. P. and Ponto, K. (2011). *More Concise Algebraic Topology*. University of Chicago Press.

Melia, J. (1999). Holes, haecceitism and two conceptions of determinism. *The British Journal for the Philosophy of Science*, 50(4):639.

Møller-Nielsen, T. (2017). Invariance, interpretation, and motivation. *Philosophy of Science*, 84(5):1253–1264.

Myrvold, W. C. (2011). Nonseparability, classical, and quantum. *The British Journal for the Philosophy of Science*, 62(2):417–432.

Nicolaescu, L. I. (1996). *Lectures on the Geometry of Manifolds*. World Scientific.

North, J. (2009). The "structure" of physics: A case study. *The Journal of Philosophy*, 106(2):57–88.

Nozick, R. (2001). *Invariances: The Structure of the Objective World*. Harvard University Press.

Pokorski, S. (2000). *Gauge Field Theories*. Cambridge University Press, 2nd edition.

Pooley, O. (2006). Points, particles, and structural realism. In Rickles, D., French, S., and Saatsi, J., editors, *The Structural Foundations of Quantum Gravity*, chapter 4, pages 83–120. Oxford University Press, Oxford.

Pooley, O. (2013). Relationist and substantivalist approaches to spacetime. In Batterman, R., editor, *The Oxford Handbook of Philosophy of Physics*, pages 522–586. Oxford University Press.

Post, H. R. (1971). Correspondence, invariance, and heuristics: in praise of conservative induction. *Studies in History and Philosophy of Science Part A*, 2(3):213–255.

Quine, W. V. O. (1961). Two dogmas of empiricism. In *From a Logical Point of View*, chapter 2, pages 20–46. Harper & Row.

Ramsey, F. P. (1931). Theories. In Braithwaite, R. B., editor, *The Foundations of Mathematics*, pages 212–236. Routledge & Kegan Paul.

Redhead, M. (1975). Symmetry in intertheory relations. *Synthese*, 32(1/2):77–112.

Redhead, M. (2001). The intelligibility of the universe. *Royal Institute of Philosophy Supplement*, 48:73–90.

Redhead, M. (2003). The interpretation of gauge symmetry. In Brading and Castellani (2003b), pages 124–139.

Riehl, E. (2014). *Categorical Homotopy Theory*. Cambridge University Press.

Riehl, E. (2016). *Category Theory in Context*. Dover.

Roberts, J. T. (2008). A puzzle about laws, symmetries and measurability. *British Journal for the Philosophy of Science*, 59:143–168.

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

Rosenstock, S. and Weatherall, J. O. (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.

Ruetsche, L. (2011). *Interpreting Quantum Theories*. Oxford University Press, Oxford.

Russell, B. (1905). On denoting. *Mind*, 56:479–493.

Saunders, S. (2013). Rethinking Newton's Principia. *Philosophy of Science*, 80(1):22–48.

Schreiber, U. (2013). Differential cohomology in a cohesive ∞-topos. arXiv:1310.7930v1.

Schreiber, U. and Waldorf, K. (2009). Parallel transport and functors. *Journal of Homotopy and Related Structures*, 4(1):187–244.

Scott, D. (1967). Existence and description in formal logic. In Schoenman, R., editor, *Bertrand Russell: Philosopher of the Century*, pages 181–200. Allen & Unwin.

Shulman, M. (2009). Homotopy limits and colimits and enriched homotopy theory. arXiv:math/0610194.

Shulman, M. (2015). Homotopy type theory: A synthetic approach to higher equalities. Forthcoming in *Categories for the Working Philosopher*, ed. Elaine Landry.

Sklar, L. (1975). Methodological conservatism. *Philosophical Review*, 84(3):374–400.

Spivak, M. (2010). *Physics for Mathematicians*. Publish or Perish, Inc.

Stanford, P. K. (2001). Refusing the devil's bargain: What kind of underdetermination should we take seriously? *Philosophy of Science*, 68(S3):S1–S12.

Strawson, P. F. (1950). On referring. *Mind*, 59:320–344.

Street, R. (1982). Two-dimensional sheaf theory. *Journal of Pure and Applied Algebra*, 23(3):251 – 270.

Swinburne, R. (1995). Thisness. *Australasian Journal of Philosophy*, 73(3):389–400.

Teh, N. J. (2016). Galileo's gauge: Understanding the empirical significance of gauge symmetry. *Philosophy of Science*, 83:93–118.

Turnbull, M. G. (2018). Underdetermination in science: What it is and why we should care. *Philosophy Compass*, 13(2):e12475.

Univalent Foundations Program (2013). *Homotopy Type Theory: Univalent Foundations of Mathematics*. http://homotopytypetheory.org/book, Institute for Advanced Study.

van Fraassen, B. C. (1970). *An Introduction to the Philosophy of Time and Space*. Random House.

van Fraassen, B. C. (1989). *Laws and Symmetry*. Oxford University Press.

Wald, R. M. (1984). *General Relativity*. University of Chicago Press.

Wallace, D. (2014). Deflating the Aharonov–Bohm effect. arXiv:1407.5073.

Wallace, D. (2017). Fundamental and emergent geometry in Newtonian physics. *The British Journal for the Philosophy of Science*. axx056.

Weatherall, J. O. (2015). Maxwell–Huygens, Newton–Cartan, and Saunders–Knox spacetimes. *Philosophy of Science*, 83(1):82–92.

Weatherall, J. O. (2016a). Fiber bundles, Yang–Mills theory, and general relativity. *Synthese*, 193(8):2389–2425.

Weatherall, J. O. (2016b). Regarding the 'hole argument'. *The British Journal for the Philosophy of Science*.

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

Weinberg, S. (1996). *The Quantum Theory of Fields*. Cambridge University Press.

Weyl, H. (1952). *Symmetry*. Indiana University Press.

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