Gauge and Explanation

Can gauge-dependent quantities be explanatory?

Caspar Jacobs

Word Count: 29,304

Submitted in partial fulfilment of
the requirements for the degree of
Bachelor of Philosophy in Philosophy

Trinity 2019

Abstract

Gauge theories, by which I mean theories with local dynamical symmetries, are ubiquitous in modern physics, extending across quantum chromodynamics, electromagnetism and (arguably) general relativity. Often philosophers also use the term 'gauge' in a broader sense as referring to those theories which contain superfluous structure or, in John Earman's memorable words, 'descriptive fluff'. In this sense, symmetries relate models which represent the same physical possibility; I call these *invariance symmetries.* It is an interesting feature of our theories that gauge symmetries are naturally interpreted as invariance symmetries.

An important debate concerns the status of quantities which are not invariant under invariance symmetries, i.e. quantities whose values differ across symmetry-related models. The orthodox position is that such quantities are not physically real, irrespective of their role in theoretical models: this is called the *Invariance Principle.* However, a question which has not received much attention is whether non-invariant quantities are *explanatory.* In other words, can interpreting such quantities as physically real increase the explanatory strength of our theories? It is surprising that the relation between invariance and explanation has received so little attention, given that an important reason for interpreting theories is to *explain* their empirically successful predictions.

In this thesis I intend to explore the relation between invariance and explanation, specifically with respect to gauge quantities. The question I set out to answer is: *can gauge-dependent quantities play an indispensable role in scientific explanations?* This investigation is pressing because if gauge quantities *are* explanatory, then there are good reasons to interpret them as physically real, contrary to the widely-held Invariance Principle.

The standard arguments for the Invariance Principle take the form of a dilemma: either symmetry-related models represent distinct possibilities, or only invariant quantities are physically real. I call the former option *absolutism.* For example, if there exist absolute positions, then models related by spacetime symmetries represent distinct possibilities. It is argued that absolutism has undesirable consequences, such as undetectability and indeterminism (e.g. the Hole Argument). This leads to the other horn of the dilemma, which necessitates the search for novel ontological posits which are invariant under symmetries. As these are relational quantities, such as relative distances, this position is called *relationism.*

I will argue that this is a false dilemma. In the course of this argument, I develop a novel approach to the interpretation of (gauge) quantities called

Quantity-Value Conventionalism. Conventionalism allows us to to interpret non-invariant quantities as physically real without being committed to considering symmetry-related models as physically distinct. It thus avoids theoretical vices such as indeterminism. Quantity-Value Conventionalism is opposed to what I call *Quantity-Value Literalism*, which is the view that all quantities have determinate values and of which both absolutism and relationism are instances.

I then argue in favour of Quantity-Value Conventionalism on the basis that realist interpretations of gauge quantities increase the explanatory strength of our theories. The general claim is that non-invariant quantities explain cosmic coincidences in the instantiation patterns of relational quantities. I develop these arguments in detail with a case study of the Aharonov-Bohm effect, in which I defend a conventionalist interpretation of the gauge-dependent vector potential field. The conclusion of this study is that the vector potential field is a physically real field which interacts with the electron wavefunction to cause the Aharonov-Bohm effect.

Acknowledgements

First and foremost, I want to thank Adam Caulton for his supervision. While writing this thesis I often vacillated between believing that what I was saying was trivially true, and that it was obviously false. Adam would convince me that sometimes it was neither, but worth exploring further. I could not have wished for a more supportive and kind supervisor, and I am looking forward to continuing working on this project together over the next few years.

The only other person to have read this thesis in its entirety is Kim. Because of your engagement I didn't have to feel like I was in it alone during the final few weeks of writing. And I promise that I will pull an actual all-nighter before finishing the DPhil.

I have also benefitted from conversations with many philosopher-friends, whose comments and support have proved invaluable: Daniel Sega Neuman, David Schroeren, James Read, Lia den Daas, Lydia Drabkin-Reiter, Natasha Oughton, Niels Martens and those who I have forgotten to mention.

I also appreciated the responses from friends from other fields: Adam Brzezinski, Ali Hazel, Itzhak Rasooly and Marco Bodnar. Where my explanations were muddled, your feedback was incisive.

Finally, thanks to the British Society for the Philosophy of Science for placing their trust in this research project by awarding me with their doctoral scholarship.

Contents

Chapter 1

The Inconsistent Triad

The following three propositions form an inconsistent triad:

(1) If a quantity is explanatory, it is physically real;

(2) No gauge quantities are physically real;

(3) Some gauge quantities are explanatory.

The Inconsistent Triad forms a philosophical puzzle which ought to be solved. In this chapter, I argue that it *is* a puzzle. I will argue that (1) follows from quite general methodological principles of realism, and that (2) is the orthodox view of gauge quantities. It follows that realists who accept the orthodoxy are committed to the denial of (3). However, there are cases in which (3) is *prima facie* true. A well-known example is the gauge-dependent vector potential in electromagnetism, which Feynman et al. (1964, Ch. 15) argues causally explains the Aharonov-Bohm effect. Recently, Rovelli (2014) has argued that gauge quantities represent "real relational structures to which the experimentalist has access in measurement".[1] There is thus support from different corners for all three propositions of the Inconsistent Triad. And hence there is a puzzle here.

I will end up arguing that realists ought to accept (3), and hence give up (2). But first I will set-up the Inconsistent Triad in more detail in this chapter. The plan is as follows. Firstly, I define three key concepts: quantities, symmetries and gauge (§1.1). Secondly, I substantiate the claim that (2) is the orthodox view of gauge quantities (§1.2). Thirdly, I argue that (1) is implied by a modest for of realism (§1.3). Finally, §1.4 sets out the plan for the remainder of this thesis.

[1] I analyse Rovelli's views in more detail in §2.4.

1.1 Background

In this section, I define the three central concepts of this thesis: quantities, symmetries and gauge. It is important to note that these are all concepts concerning our *representations* of the world. For example, I will think of quantities as *functions*, which are causally inert mathematical objects; but some quantities *represent* properties and relations which exist in nature and which causally interact with each other.

On the *semantic view* of theories, a theory is presented as a family of *models* (Suppes 1960). We can take these models to be tuples of mathematical objects $\langle A, B, C, ...\rangle$ which are usually required to satisfy one or more equations or axioms. For example, the content of Newtonian Gravitation is expressed through models of the form $\langle M, t_{ab}, h^{ab}, \nabla, \sigma^a, \phi, \rho\rangle$: M is a four-dimensional smooth manifold; t_{ab} and h^{ab} are compatible temporal and spatial metrics ($t_{an}h^{nb} = 0$); ∇ is an operator encoding a standard of straight—or uniform—motion compatible with both metrics ($\nabla_a t_{bc} = \nabla_a h^{bc} = 0$); σ is a time-like vector field identifying points of space over time; and ϕ and ρ are scalar fields corresponding to the gravitational field and the matter distribution. These are the *kinematically possible models* (KPM) of Newtonian Gravitation. If a model also satisfies the following equations, it is a *dynamically possible model* (DPM):

$$R^a_{bcd} = 0 \tag{1.1}$$

$$h^{ab}\nabla_a\nabla_b\phi = 4\pi\rho \tag{1.2}$$

where $R^a_{bcn}\xi^n = \nabla_{[b}\nabla_{c]}\xi^a$. Here, (1.1) imposes flatness on ∇ and (1.2) is the Newton-Poisson equation which fixes the gravitational field given a distribution of matter.

Call a theory's KPMs its *states*. The states represent kinematically possible *worlds*, which we can think of as those worlds which have the right 'form' for a particular theory: world which contain the right kinds of objects and properties set on an appropriate spacetime. A subset of states represents the dynamically possible worlds: those worlds which are 'allowed' to obtain if the theory is true. For example, the dynamically possible worlds of Newtonian Gravitation are represented by those models which satisfy equations (1.1) – (1.2). In other theories, states are instead taken to represent *instantaneous* states of the world, with the dynamical equations specifying the allowed *trajectories* through state space, but nothing in the present discussion hangs on this difference.

1.1.1 Quantities

We can understand quantities as *functions* from the set of states, $\mathcal{S}$, into a *value space*, $\mathcal{V}$, which is a mathematical space with a particular structure.[2] For example, if a quantity Q can take on any positive value, then its value space $\mathcal{V}_Q$ is isomorphic to the positive real numbers, $\mathbb{R}^+$. Q is then a function from the set of states into the positive reals, i.e.

$$Q : \mathcal{S} \to \mathbb{R}^+. \tag{1.3}$$

The value of a quantity in a given state s, which we write $Q(s)$, is thus a particular element of its associated value space, such as a particular positive real number. Of course, values need not be numbers. For example, the value space of (absolute) velocity is a three-dimensional real vector space.

Neither the quantities themselves, which are functions, nor their associated value spaces are elements of physical reality. Both are mathematical objects which are part of our theories. I am agnostic as to whether such mathematical objects are also 'real'. However, we sometimes want to *use* quantities to represent something else which *is* real in the physical sense, such as the mass of the Great Sphinx. I will call a quantity which represents a physical entity (such as an individual, a property or a relation) a *physically real quantity*. The values of a physically real quantity then represent (physical) *magnitudes*, such as '180 cm' or '65 kg'. For example, 'height space' is a space isomorphic to $\mathbb{R}^+$ whose values refer to height magnitudes like '170 cm', '180 cm', '2 m', '1 inch'. Note that a (dimensionful) value space is not tied to a particular unit, such as metres; I take it that, for example, '170 cm' and '1.7 m' are different names for the *same* magnitude.

Generally, not all quantities are physically real, and hence not all values refer to physical magnitudes. For example, if I define a function which maps each state onto the number three, there is no reason to believe that this function represents something real. I will pick up the question of when we *should* believe that a quantity is physically real in §1.3.

1.1.2 Symmetries

The notion of symmetry is commonly applied to particular objects, for example when I say that my left and right shoe are mirror symmetrical. The symmetries we are concerned with, on the other hand, pertain to theories

[2] The idea of a value space for properties goes at least back to Van Fraassen (1967), and similar ideas have recently been defended by Denby (2001), Funkhouser (2006) and Caulton (2015).

rather than objects; like quantities, symmetries are elements of our representations.

How to define the symmetries of a theory has been discussed in much detail (see, for example, Ismael and Van Fraassen (2003) and Belot (2013)). Therefore, I will be brief here. The basic notion of symmetries is that of a transformation on the elements which occur in the kinematically possible models, such as t_{ab} or ρ in the models above. Since we have identified states with KPMs, symmetries are mappings from the set of states onto itself.

Ismael and Van Fraassen (2003) have argued that we can use symmetries 'as a guide to superfluous structure'. This is usually understood as the claim that "if two mathematical models of a physical theory are related by a symmetry transformation, then those models represent one and the same physical state of affairs" (Greaves and Wallace 2014, 60). In other words, symmetry-related models are *physically equivalent*. Call symmetries which relate physically equivalent models *invariance symmetries*.

As Belot (2013) points out, not all symmetries are invariance symmetries. Consider for example the proposal that the invariance symmetries are those symmetries which map *dynamically* possible models onto each other. Here is a counter-example to this proposal. Recall Galileo's famous ship: a constant velocity boost of the ship and its contents is a symmetry of Newtonian mechanics, since adding a constant term to the ship's velocity generates another DPM. Nevertheless, a velocity boost clearly induces an *empirical* difference, since we can measure the velocity of the ship relative to the shore. Therefore, two models which differ by a local velocity boost are naturally taken as representing *distinct* states of affairs, and so not all dynamical symmetries are invariance symmetries.

For the purposes of this thesis it does not matter if we can generally characterise the symmetries which relate physically equivalent states. I will simply use the term 'invariance symmetries' as standing for whatever symmetries play this role. As it is often clear from the context what is meant, I will often refer to invariance symmetries as 'the symmetries of our theories' or simply 'symmetries'.

1.1.3 Gauge

Finally, what is gauge? In this thesis, I will speak of gauge theories; gauge symmetries; and gauge quantities. The first, gauge theories, are theories which possess a gauge symmetry. And the third, gauge quantities, are those quantities whose values vary under gauge symmetries, i.e. quantities which have different values across models related by gauge transformation. The crucial notion, then, is that of a *gauge symmetry*.

Unfortunately, there are multiple definitions of gauge symmetries in circulation. Weatherall (2015) helpfully distinguishes between two usages of the term. On the first usage, which is common among philosophers, gauge theories exhibit superfluous structure in the sense discussed above. On this definition invariance symmetries *are* gauge symmetries. On the second usage, on the other hand, gauge symmetries are defined by their mathematical properties. For example, Trautman (1980, 26) defines gauge theories as theories "of a dynamic variable which, at the classical level, may be identified with a connection on a principal bundle"; similarly, Kosso (2000) uses the term 'gauge symmetries' to refer to *internal* and *local dynamical* symmetries, where dynamical symmetries are those symmetries which map dynamically possible models onto each other.

The question under discussion in this thesis is whether gauge quantities are physically real. If gauge symmetries encode excess structure, it is a small step to denounce gauge quantities as part of this excess. In order not to prejudice the debate, I will therefore use 'gauge' in Weatherall's second sense. Specifically, I take gauge symmetries to be local symmetries, i.e. dynamical symmetries which depend on space-time coordinates. A prime example is the gauge freedom of semi-classical electrodynamics, which allows space-time dependent transformations of the vector potential field.[3]

Whether gauge symmetries in this sense are invariance symmetries is a highly non-trivial question. Caulton, using the term 'analytic symmetries' for something much like invariance symmetries, writes that:

> It may indeed be claimed that all gauge symmetries are analytic and all analytic symmetries are gauge. But establishing that claim would be a significant philosophical achievement [...]. Besides, it seems to me that there are both analytic symmetries that are not gauge and gauge symmetries that are not analytic. (Caulton 2015, 157)

I agree with Caulton that it is in principle possible for gauge symmetries to relate empirically distinguishable states. For a toy example, consider a

[3] I don't require that gauge symmetries are *internal*, i.e. apply only to non-spatiotemporal degrees of freedom. I understand Wallace (2015) as arguing that we can reformulate external symmetries, such as those of general relativity, as internal symmetries, in which case the distinction is not fundamental. However, not much depends on this choice.

field much like the vector potential field in classical electrodynamics, that is, a field O^μ satisfying:

$$F^{\mu\nu} = \partial^\mu O^\nu - \partial^\nu O^\mu \tag{1.4}$$

where $F^{\mu\nu}$ is a given field tensor. The following transformation is a symmetry of this theory:

$$O^\mu \to O'^\mu = O^\mu - \partial^\mu f(x), \tag{1.5}$$

where $f(x)$ is an arbitrary function. (1.5) is a local transformation, since it depends on x, and hence this is a gauge theory.

Now suppose that unlike the electromagnetic vector potential, O^μ is observable. For example, suppose that our noses contain small tendrils which dynamically interact with the O^μ-field. Then it will be possible to empirically distinguish between O and O', despite the fact that these solutions are related by a gauge symmetry. Hence, the gauge symmetry (1.6) relates physically distinct states, and is thus not an invariance symmetry. This example is analogous to Galileo's ship considered above, since in both cases a (dynamical) symmetry relates empirically inequivalent states of affairs.

Thus, gauge symmetries are not necessarily invariance symmetries. Nevertheless, most interesting cases of gauge symmetries in physics *are* naturally understood as invariance symmetries: examples range from the gauge freedom in semi-classical electrodynamics to the diffeomorphism invariance of general relativity, as well as the $U(1) \times SU(2) \times SU(3)$ symmetry group of the Standard Model. These symmetries all relate states which are in principle indistinguishable and thus are naturally understood as physically equivalent. I am not aware of any non-contrived example of gauge symmetries which relate empirically distinguishable states. Indeed, the question of *why* this is so is an interesting one — why *don't* we have gauge theories of observable quantities? I do not have an answer to this question, and in the remainder of this thesis I will assume that all gauge symmetries are invariance symmetries. Nevertheless, it is worth keeping in mind that, for all we know, this is a contingent fact about the physical theories which best describe *our* world, rather than a conceptually necessary fact about gauge.

1.2 The Invariance Principle

Recall that (2) states that *gauge dependent quantities are not physically real.* The first claim of this chapter is that (2) of the Inconsistent Triad represents the orthodox view about gauge quantities. (2) follows from what is often called the *Invariance Principle*, which is the broader claim that "only quantities that are invariant under the symmetries of our theories are physically real" (Møller-Nielsen 2017, 1253). 'The symmetries of our theories' here refers to what I call invariance symmetries: those symmetries which relate physically equivalent models. The Invariance Principle states that only those quantities whose values do not change under such symmetries are physically real. In other words, if we have a set of symmetry-related states, then only those quantities which map all these states to the same value in value space are physically real.

The Invariance Principle has gathered assent from a broad range of eminent physicists and philosophers throughout the 20^{th} and 21^{st} century. Here is a collection of representative quotes from two philosophers and a physicist:

> "Physically real quantities are invariant under exact symmetries - this is the general lesson." (Saunders 2003, 300)

> "[O]nce we possess the covariant representation under which the equations stay the same for all coordinate systems, the quantities in the (covariant) equations are the real and objective quantities." (Nozick 2001, 82)[4]

> "[T]he important things in the world appear as the invariants [...] of these transformations." (Dirac 1930, vii)[5]

Endorsements of the Invariance Principle can also be found in Saunders (2007), Baker (2010), Dewar (2015), Caulton (2015) and Dasgupta (2016), among others. Clearly, if (2) is implied by the Invariance Principle then it, too, is part of the orthodox view.

Is (2) implied by the Invariance Principle? This depends on the answer to the question we discussed above: whether gauge symmetries are also invariance symmetries. If they are, then gauge symmetries fall under the

[4]Note that Nozick's take on the Invariance Principle is subtly different: not the quantities which are invariant under symmetry transformations, but those which occur in the *equations* which are invariant under symmetry transformations are physically real. On this criterion, a non-invariant quantity like the vector potential field *is* physically real, since it occurs in the Hamiltonian of a charged particle moving through a magnetic field. As we will see in the later chapters, I believe this sort of approach is on the right track.

[5]The phrase 'important things' is sufficiently vague to admit of multiple interpretations, but I believe it is naturally understood as meaning 'fundamental quantities'.

purview of the Invariance Principle. As I already mentioned, the interesting instances of gauge quantities concern unobservables and so the Invariance Principle applies to them. For example, the electromagnetic vector potential is an unobservable field which varies under the gauge transformations of semi-classical electrodynamics. Therefore, for our purposes we can consider (2) as a special case of the Invariance Principle, and I will do so for the remainder of this thesis.

1.3 Explanation

We now come to (1) of the Inconsistent Triad: explanatory quantities are physically real. I argue that (1) follows from broad realist commitments in the philosophy of science. This realist commitment is roughly that we need to posit unobservable entities in order to explain our scientific observations. I will elaborate on this below.

The debate between realists and anti-realists is about how to *interpret* our theories. Firstly, note that the theories we are talking about are already *partially interpreted*: specifically, most of their observational content is already determined. Otherwise, we would not know whether our theories are empirically successful (indeed, we would not even know if their subject matter was physics or biology!), and so there would be nothing for the realist to explain.[6] What the debate is about, then, are those quantities which putatively refer to *unobservable* entities. Interpreting these quantities literally means interpreting them as directly referring to unobservable entities. Compare this with the anti-realist (of a positivist stripe), who interprets unobservable quantities *non-literally* as indirectly referring to observables, for example via linguistic bridging principles.

The realist posits unobservable quantities because she believes their existence is indispensable in explaining the successful predictions of our theories. This, for example, is the point of Putnam's (1975, 73) famous claim that "the positive argument for realism is that it is the only philosophy that doesn't make the success of science a miracle." But we do not have to buy into full-blown scientific realism in order to accept this conclusion. A Quine-inspired empiricist, for example, believes that we are committed to the existence of those quantities which are indispensable in organising and explaining our experiences (the "irritations of our sensory surfaces"; (Quine 1966)). Therefore, the notion of explanation is important whenever we are interpreting the unobservable parts of our theories, and the realist's best

[6] I take these observations about interpretation from Ruetsche (2011, Ch. 1).

argument for the existence of unobservable quantities is their contribution to the explanatory strength of our theories.

Of course this leaves it unclear what exactly explanation *is.* I am not going to define 'explanation', since I believe that it is a pluralistic concept: there are causal explanations, nomological deductions, unifications and other sorts of explanations. Hence, different physical posits can increase the explanatory strength of our theories in different ways. If we try to give a general characterisation of scientific explanation, we may overlook novel forms of explanation. For example, Weatherall (2011) details a type of explanation which does not easily fit into any existing category. I will therefore leave 'explanatory' as a placeholder term to be filled in on a case by case basis. In Chapter 4 I will detail the specific sense in which the non-invariant quantities we are interested in are explanatory.

In any case, it is a bad idea for the realist to interpret *all* quantities as physically real. In a seminal paper, Larry Laudan (1981) mounts an attack against realism by presenting a range of examples of successful theories whose central terms are now taken not to refer. If these theories were successful despite their non-referring terms, Laudan argued, then clearly it is not the existence of these unobservable entities which explains their success.

The 'explanationist' answer to Laudan's challenge is that the non-referring quantities which Laudan points out were inessential to the theories in question: they are 'idle posits' as opposed to 'working parts'.[7] According to the explanationist, the realist is only committed to the existence of those quantities which actually do explanatory and predictive work, that is, those quantities which are indispensably responsible for explaining the theory's success. Psillos, for example, writes that "the theoretical constituents whose truth-likeness can best explain empirical successes are precisely those which are essentially and ineliminably involved in the generation of the predictions and the design of the methodology which brought these predictions about" (Psillos 1999, 78). And Harker (2010, 195) argues that "[n]ew theories recommend themselves, in part, by explaining what available alternative theories cannot". This means that it is the explanatory strength of individual quantities with which the realist should be concerned.

Some *non-explanatory* quantities may nevertheless remain indispensable for characterising theoretical models. For example, imaginary numbers are indispensible to quantum mechanics, but we do not have to reify them to explain quantum mechanics' successful predictions. Dewar summarises this caveat well: "if [quantities] are only indispensable for the purposes of char-

[7]This and similar distinctions are found in Kitcher (1993); Psillos (1999); Saatsi (2005)

acterising or representing the structures that are (explanatorily, hence physically) significant, then it seems that we have much greater leeway to regard them as mere mathematical scaffolding, without being honour-bound to accord them physical significance" (Dewar 2015, 322). There is thus a difference between quantities which are explanatory in the sense that we use them to *give* scientific explanations, and quantities which are explanatory in the sense that *positing* such quantities as physically real increases the explanatory strength of our theories. We are concerned with the latter.

In conclusion, the realist should take *explanation* as her guide to superfluous structure: quantities which are explanatory are physically real. As a result, realism is committed to (1) of the Inconsistent Triad.

1.4 Conclusion

I have defined quantities, symmetries and gauge in §1.1, arguing that these are all concepts which pertain to our representations of reality. In §1.2, I showed that (2) of the Inconsistent Triad is an implication of the orthodox Invariance Principle. Finally, §1.3 argued that realism is committed to (1) of the Inconsistent Triad.

Given that there are cases in which (3) is *prima facie* true, it follows that the Inconsistent Triad presents a puzzle for the realist. In the remaining chapters I will attempt to solve this puzzle, arguing that some gauge quantities *are* explanatory, and hence that we ought to reject (2) in favour of (3). I will now briefly outline the structure of the remaining chapters.

Chapter 2 and 3 together form an attack on the Invariance Principle, and thus on (2) of the Inconsistent Triad. In Chapter 2, I present two approaches to the interpretation of quantities, which I call Quantity-Value Literalism and Quantity-Value Conventionalism. The former encompasses both absolutism and relationism. I will show that the latter present us with an intermediate position which has more content than relationism but less content than absolutism. In Chapter 3, I argue that if we accept Quantity-Value Conventionalism we need not posit the Invariance Principle in order to avoid theoretical vices such as undetectability and indeterminism. This means that we can interpret non-invariant quantities as physically real without incurring any of the associated costs.

In Chapter 4, I argue that there are good reasons for doing so: non-invariant quantities can increase the explanatory strength of our theories. Specifically, I will argue that we must posit such quantities in order to explain what are otherwise 'cosmic coincidences'. This completes the attack against the Invariance Principle: it is both possible and desirable to interpret non-invariant quantities as physically real.

Finally, Chapters 5 and 6 consider Quantity-Value Conventionalism in the context of gauge theories. In Chapter 5, I give an overview of the Aharonov-Bohm effect and its interpretations. These interpretations are either absolutist or relationist (and thus in both cases literalist). I will argue that these options either lead to indeterminism or to cosmic coincidences. In Chapter 6, I show that a Conventionalist alternative allows us to explain the coincidences while avoiding indeterminism. The resulting interpretation is realist about the vector potential field, and so I conclude that (2) is false: there are physically real gauge quantities.

Chapter 2

Quantities: Between Absolutism and Relationism

According to proposition (1) of the Inconsistent Triad, explanatory quantities are physically real. And according to (3), some gauge quantities are explanatory. It follows that some gauge quantities are physically real: the negation of (2). So if we have an argument that gauge quantities *are* explanatory, we can infer from modus ponens that (2) is false under scientific realism. However, one person's modus ponus is another's modus tollens, and the orthodox view has it that since (2) is true, gauge quantities *cannot* offer genuine explanations. The result is a standoff.

Therefore, I proceed differently. In this and the next chapter, I argue directly against the Invariance Principle, which underpins (2). If the argument succeeds, we are free to assess the reality of gauge quantities on the basis of their explanatory merit, without an *a priori* presumption against their reality. This approach is consistent with the emphasis on explanation I defended in the previous chapter.

The dialectic of the following two chapters is as follows. In this chapter, I outline two alternative approaches to the interpretation of quantities, which I call Quantity-Value *Literalism* (§2.1), and Quantity-Value *Conventionalism* (§2.2). According to the former position, the values of quantities consistently refer to particular physical magnitudes across theoretical models. On the latter view, on the other hand, it is partly conventional which dimensionful value we use to refer to particular magnitudes. The remainder of this chapter works out the details of Quantity-Value Conventionalism: §2.3 gives it a perspicuous metaphysical characterisation in the form of *anti-quidditism* about magnitudes; §2.4 discusses the sense in which conventionalism reveals the relational structure of the world; and §2.5 considers the role of symmetries on the conventionalist account.

In Chapter 3 and 4, I relate this to the Inconsistent Triad. In Chapter 3, I argue that arguments for the Invariance Principle, and hence for (2), implicitly assume Quantity-Value Literalism. If we instead accept Quantity-Value Conventionalism, the Invariance Principle does not follow. In other words, with Quantity-Value Conventionalism we can consider gauge quantities as physically real. I then argue in Chapter 4 that doing so increases the explanatory strength of our theories. Therefore the realist ought to favour a conventionalist-realist interpretation of gauge quantities, affirming (3).

2.1 Quantity-Value Literalism

The following principle regarding quantities and their values is widely believed:

> *Quantity-Value Literalism*: Quantities are physically real only if their values consistently refer to the same physical magnitudes across theoretical models.

This requires some unpacking. As we have seen in Chapter 1, generally only *some* mathematical quantities represent physical objects, properties, relations, et cetera. These are the *physically real* quantities. Quantity-Value Literalism (QVL) places a constraint on which quantities we are allowed to interpret as physically real, namely only those whose values we can interpret as denoting the same magnitudes across models. QVL is thus first and foremost a principle regarding the interpretation of theories, rather than a metaphysical rule — though as we will see below, it is not metaphysically neutral.[1]

Recall from the previous chapter that to each quantity there is associated a particular (mathematical) value space. If a quantity is physically real then its values — the elements of its value space — refer to physical magnitudes, such as mass properties or lengths. What QVL demands is that the individual values of a quantity consistently refer to the *same* physical magnitude across theoretical models. For example, the mass value '5 kg' must consistently refer to the property of having a mass of five kilograms; if we have two models such that some object is assigned the mass value '5 kg'

[1]Specifically, QVL is not the same as the metaphysical thesis of *quidditism*. I discuss quidditism in §2.3.

in both, then these models represent two worlds in which that object has the *same* mass. I will call quantities whose values consistently refer to the same physical magnitude across theoretical models *determinate quantities*.

QVL has the following consequence: if we interpret a quantity as physically real, then the models across which its values vary are physically inequivalent. Consider for example absolute position in Newtonian mechanics.[2] In models of Newtonian mechanics, the quantity $x_B(t)$ represents the spatial location of body B at time t; it is assigned a value from a three-dimensional Euclidean 'location space'. According to QVL, these values denote particular points of spacetime across physical models. Therefore, two models which assign different values to $x_B(t)$ represent B as located at *different* locations, and hence are physically inequivalent.

Conversely, if we interpret distinct models as physically equivalent, then QVL demands anti-realism about quantities whose values differ across these models. Consider for example the so-called *static shift*: uniformly displacing the entire material universe by a constant amount, say two metres north. If we consider static shifts as invariance symmetries of Newtonian mechanics, then shift-related models are physically equivalent. In that case, QVL implies that position is *not* a physically real quantity. For consider the following two shift-related models:

$$\mathcal{M}_1 : \begin{cases} x_A = (1,0,0) \\ x_B = (2,0,0) \end{cases} \qquad \mathcal{M}_2 : \begin{cases} x_A = (2,0,0) \\ x_B = (3,0,0) \end{cases}$$

If these models are physically equivalent, then $(1,0,0)$ and $(2,0,0)$ both represent the same spacetime point: A's location. Similarly, $(2,0,0)$ and $(3,0,0)$ both represent B's location. This means that the same value, $(2,0,0)$, represents A's location in $\mathcal{M}_1$, and B's location in $\mathcal{M}_2$. But both models represent these bodies as occupying the *distinct* locations. It follows that $x(t)$ is *not* determinate: its values refer to different spatial locations across theoretical models.

If the literalist nevertheless wants to identify shift-related models as physically equivalent, she can instead interpret *distances* as the fundamental physically real quantities. We can define the distance between a and b as $d_{ab} = |x_a - x_b|$, which is invariant under static shifts. Call this position *relationism*. QVL then implies that distances are determinate. Of course, it is possible that we discover that distance also varies under symmetries, such as uniform spatial scaling transformations. The literalist then again

[2] With 'absolute' I mean that position is a *monadic* property. See Martens (2019b, fn. 4).

faces the choice between keeping distances and distinguishing scaling-related models, or identifying scaling-related models and adopting an ontology of relations *between* distances. But in each case literalism is not *committed* to one of these choices; it all depends on which symmetries she regards as invariance symmetries. Therefore, QVL itself is *neutral* between absolutism and relationism.

Finally, I offer a remark on what is *literalist* about Quantity-Value Literalism. According to QVL, models which differ in the values of physically real quantities invariably represent distinct possibilities. Therefore, values are 'literally' interpreted as standing in a one-to-one representation relation with particular magnitudes. This usage of 'literal' is widespread. Read and Møller-Nielsen (2018, 8), for example, claim that "straightforwardly—or 'literally'—understood, the world represented by [$\mathcal{M}_1$] differs from that represented by [$\mathcal{M}_2$]", and similar claims are found in Dewar (2018, 253) and Møller-Nielsen (2017, 1260).

2.2 Quantity-Value Conventionalism

There is a reasonable alternative to Quantity-Value Literalism, which goes back to Stalnaker (1979, 352-4). Stalnaker's proposal there is that the values of (some) physically real quantities are (sometimes) *conventional*. I call this alternative position *Quantity-Value Conventionalism* (QVC):

> *Quantity-Value Conventionalism*: Quantities which are physically real may have values which denote different physical magnitudes across theoretical models, dependent on conventional choices.

QVL and QVC are mutually exclusive alternatives. According to the former, quantities are physically real only if their values refer to the same magnitudes across models, whereas the conventionalist believes that distinct models may use the same value to refer to different magnitudes.

The view which Stalnaker presents is more specific than Quantity-Value Conventionalism as I have stated it. His proposal amounts to the following claim:

> *Meaningfulness Criterion*: A statement is (potentially) *physically meaningful* if and only if it is invariant across physically

equivalent models.[3]

From left to right, the Meaningfulness Criterion states that a statement is physically meaningful only if it is invariant across physically equivalent models. This is uncontroversial: physically equivalent models agree on all physical facts, so whatever varies across such models cannot have a consistent meaningful interpretation.

But read from right to left, the Meaningfulness Criterion is highly non-trivial: it states that *any* invariant statement is potentially physically meaningful, *even if* such a statement involves non-invariant quantities. We will see examples of such statements below. The disclaimer 'potentially' indicates that we are not committed to interpreting *all* invariant sentences as meaningful. But we are allowed to interpret *any* invariant sentence as meaningful if there is good reason for doing so. The Quantity-Value Conventionalist endorses the Meaningfulness Criterion.

It is because of the Meaningfulness Criterion that QVC is different from both (literalist) absolutism and (literalist) relationism. Suppose again that static shifts are invariance symmetries. Then ascriptions of determinate locations are not meaningful according to the Meaningfulness Criterion, since such ascriptions are not invariant under static shifts. So QVC has less content than the literalist form of absolutism. But now consider a statement such as "particle b has *some* location". The relationist denies this, since she believes that there are *no* locations. According to the Meaningfulness Criterion, on the other hand, this statement *is* meaningful; if particle b has a location in one model, then it has a location in any other model related to the first by a static shift. Thus, QVC has *more* content than relationism. There is a 'truth gap' between absolutism and relationism which conventionalism fills.

We can understand the Meaningfulness Criterion as demanding *supervaluation* over physically equivalent models: any statement which is true (false) in all symmetry-related models is *supertrue* (superfalse). QVC implies realism about supertrue and superfalse statements. Interestingly, supervaluation does not commute with logical composition. A different explanation of the difference between literalism and conventionalism then is that the former *first* supervaluates over 'atomic' theoretical statements and then composes more complex sentences, whereas the conventionalist first composes complex sentences and *then* supervaluates.

Let $[\phi(s)]_{\mathcal{I}}$ denote the truth value of a sentence ϕ in a model s when we

[3]Cf. Mundy (1986).

supervaluate it with respect to a set of invariance symmetries $\mathcal{I}$.[4] The truth conditions of $[\phi(s)]_{\mathcal{I}}$ are:

$$[\phi(s)]_{\mathcal{I}} = \begin{cases} 1 \text{ iff } \phi(i(s)) = 1 \text{ for all } i \in \mathcal{I} \\ 0 \text{ iff } \phi(i(s)) = 0 \text{ for all } i \in \mathcal{I} \\ \# \text{ otherwise} \end{cases} \tag{2.1}$$

Now consider the statement that a quantity Q has *some* value in s: $\exists(x)\, Q(s) = x$. We can write this as the following disjunction:[5]

$$Q(s) = x_1 \vee ... \vee Q(s) = x_n \tag{2.2}$$

The crucial insight here is that there is a difference in truth value between the supervaluation of this sentence as a whole and the composition of its individually supervaluated disjuncts (i.e. supervaluation and composition do not commute). In other words,

$$[Q(s) = x_1 \vee ... \vee Q(s) = x_n]_{\mathcal{I}} \not\Leftrightarrow [Q(s) = x_1]_{\mathcal{I}} \vee ... \vee [Q(s) = x_n]_{\mathcal{I}} \tag{2.3}$$

Proof. If Q is not invariant under $\mathcal{I}$, then $[Q(s) = x]_{\mathcal{I}}$ is neither true nor false (for any x), since the transformations in $\mathcal{I}$ change Q's value. Under usual trivalent semantics (e.g. Kleene), the disjunction of sentences which lack bivalent truth values itself also lacks a bivalent truth-value. Therefore, (2.2) is neither true nor false. This is of course exactly what QVL implies: if quantities are not invariant under invariance symmetries, then there are no meaningful statements about them. On the other hand, $[Q(s) = x_1 \vee ... \vee Q(s) = x_n]_{\mathcal{I}}$ *does* have a truth value, since on any $\mathcal{I}$-related model Q has *some* value between x_1 and x_n. And this is in accordance with the Meaningfulness Criterion, which states that *any* sentence which is invariant across models is meaningful. □

We can put the difference between QVL and QVC as follows. On QVL, we first gather the atomic sentences of our theory — individual value ascriptions — and supervaluate over these. We can then compose physically meaningful sentences out of the atomic sentences which are either supertrue or superfalse, but no meaningful statements out of those which lack a truth value. On QVC, on the other hand, we first compose any complex sentence

[4] Compare the role of *interpretations* in logic.

[5] Of course, most realistic quantities are continuous, so the disjunction would have an infinite number of disjuncts. I will ignore any complication which could arise from this.

we need out of the simple ones. We *then* supervaluate over these sentences, and any sentence which is supertrue or superfalse is (potentially) physically meaningful. Since composition and supervaluation do not commute, these procedures have different outcomes. And this allows conventionalism to fill the truth gap which literalism leaves between absolutism and relationism.

In the next two subsections I illustrate Quantity-Value Conventionalism with two examples. The first is an example of spatial which I take from Stalnaker. The second is a novel application of Stalnaker's approach to non-spatial quantities such as mass. In the course of explaining these it will become clear what is *conventionalist* about Quantity-Value Conventionalism.

2.2.1 Space

Suppose that models related by uniform global rotations are physically equivalent, and hence that there are no absolute positions (the argument is analogous to the one from static shifts). The following three diagrams then represent the same state of affairs:

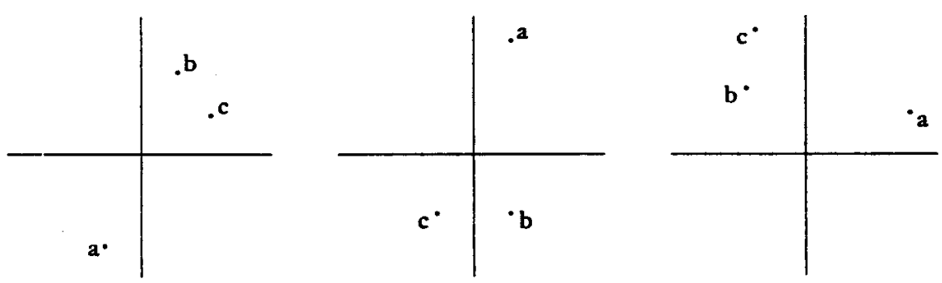

Figure 2.1: Reproduced from Stalnaker (1979).

Here, the dots represent the locations of three bodies labeled 'a', 'b' and 'c' in two-dimensional position space. The diagrams agree on all distances between bodies, and hence are physically equivalent models. The values of the locations of each dot, on the other hand, are different in each diagram, so the Meaningfulness Criterion implies that claims about them are not physically meaningful.

If we accept QVL, then this implies that spatial location itself is not a physically real quantity. Again, the argument mirrors the example using static shifts given above. But according to conventionalist there *are* physically meaningful statements about spatial position. Consider the claim that particle b has *some* location in space (i.e. is assigned *some* value from two-dimensional position space). This statement is invariant under shifts and rotations: b is assigned a location on all three diagrams in Fig. 2.1. From the Meaningfulness Criterion it follows that this statement is (potentially) physically meaningful. And the same goes for other invariant statements, such as "the location of b is closer to c than to a". Stalnaker summarises the point as follows: "in the relational theory of space, spelled out in this way, no intrinsic spatial properties would be real (*except the property of being located somewhere in space*)" (Stalnaker 1979, 353-4, emphasis mine).

Stalnaker claims that on this view "logical space itself is in part conventional" (353). It is conventional in the sense that it is up to us which of the three diagrams in Fig 2.1 we use to represent the locations of a, b and c. This is not so for the literalist, who considers these diagrams as representing distinct possibilities (or as not representing physically real quantities at all). This means that for the conventionalist there are elements of location space, such as its overall orientation, which are *not* representational. Therefore, the position space of the conventionalist is different from the position space of the literalist absolutist, since the former has no distinguished zero element and no distinguished orientation. In general, then, the value spaces of the conventionalist have *less* structure than those of the absolutist. At the same time, conventionalism has *more* structure than relationism, since the latter advocates adopting a smaller invariant value space such as distance space instead. We can see from this that conventionalism represents an option in between (literalist) absolutism and relationism.

Some caution is needed, since there is another sense in which representation is conventional for *both* the literalist and the conventionalist. On either view, we make a conventional *choice of units*. For example, which object represents the 'standard kilogram', or what our spatial axes are. A change of units is what is known as a *passive transformation* and has no physical significance. But on the literalist view there is no more room for convention once our choice of units is fixed. For example, once we have chosen our spatial axes the diagrams in Fig. 2.1 *must* represent different possibilities (if space is real). But the conventionalist believes that there are conventional elements to our representations *even after we have fixed our coordinates*, such as rotations in location space with respect to chosen axes. The conventionalist thus believes that our representations are conventional to a greater degree than the literalist.

2.2.2 Properties

Stalnaker applies his conventionalism to spatial properties, but in principle the Meaningfulness Criterion also concerns quantities such as mass, charge and field strengths. I will now give a brief outline of how this goes in the case of mass.

It is sometimes claimed that Newtonian mechanics is invariant under *mass scalings*: if we multiply all mass values by a constant, we obtain an empirically equivalent dynamically possible model of Newtonian mechanics.[6] If we consider scaled models as physically equivalent, then QVL implies that mass is physically unreal. Consider the following two models:

$$\mathcal{M}_1 : \begin{cases} m_A = 5 \text{ kg} \\ m_B = 10 \text{ kg} \end{cases} \qquad \mathcal{M}_2 : \begin{cases} m_A = 10 \text{ kg} \\ m_B = 20 \text{ kg} \end{cases}$$

The argument is analogous to the one given in §2.1. Since these models are physically equivalent, both represent m_B as having the same mass. In $\mathcal{M}_1$, this mass is referred to as '10 kg', but in $\mathcal{M}_2$ the same value refers to the distinct mass of A. Therefore, mass values do not consistently refer to the same mass magnitudes and so mass is not a determinate quantity.

Again, a possible response to this argument is to take up relationism and posit that mass *ratios* are fundamental. But QVC provides another alternative. In Chapter 4 I will argue that this alternative is superior; here, I simply point out that it is a possible route one can take.

According to QVC there *are* meaningful statements about masses which are not captured by relationism. For example, it is invariant under mass scalings that all bodies have *some* mass, i.e. that all bodies are assigned some mass value. And *laws* about mass are also invariant; for example, it is true across models that the gravitational force with which a massive body attracts other massive bodies is proportional to its own mass. The conventionalist can thus interpret mass as a physically real quantity which figures in our laws. On the other hand, the conventionalist does not believe that scaling-related models are physically equivalent, and so differs from the (literalist) absolutist in that she does not posit undetectable intrinsic mass values.

We can give a similar account for other non-invariant quantities, with the caveat that different quantities vary under different symmetries. For

[6] Whether mass scalings *are* symmetries of Newtonian mechanics is a matter of dispute; see Dasgupta (2013) for the positive case, and Martens (2019b) for a counterargument. Since I am only using mass as an example, the outcome of this particular debate does not matter here.

example, it is often claimed that thermodynamics is invariant under linear transformations of temperature.[7] If this is true, then we can meaningfully talk about ratios of temperature differences, but not of ratios of temperatures. The account also applies to *gauge* quantities, but with some interesting differences. Since gauge symmetries are local symmetries, QVC implies that even values within the same model do not consistently refer to the same magnitude if they are evaluated at different spacetime points. I will discuss this novel implication in detail in Chapter 6.

2.3 Anti-Quidditism

According to the so-called *motivationalist* approach to symmetries, we are only allowed to interpret symmetry-related models as physically equivalent if we can explain their equivalence in terms of an underlying metaphysical picture. Otherwise, "the reality in terms of which this physical equivalence is to be understood will [...] remain opaque" Møller-Nielsen (2017, 1263). The motivationalist demands a "perspicuous metaphysical picture" which underpins our interpretative choices.

I am sympathetic to motivationalism, and in this section I aim to offer a perspicuous metaphysical picture in terms of which we can understand QVC. In fact, I believe we already have such a picture in the case of QVC as applied to spatial location: *sophisticated substantivalism*.[8] According to sophisticated substantivalism, spacetime is a physically real substance in which bodies are located but shift-related models are physically equivalent. In other words: unlike relationists, sophisticated substantivalists believe that bodies have *a* location in space, but unlike absolutism this location is not *determinate*, since we cannot use coordinates to refer to the same location across models.

But whereas conventionalism is in the first place a position regarding the interpretation of quantities, sophisticated substantivalism comes with an accompanying metaphysical picture of spacetime: *anti-haecceitism*. According

[7]Nozick (2001, 77) claims that the invariance of temperature under linear transformations is reflected in the fact that the Celsius, Fahrenheit and Kelvin scales are related by linear transformations. Skow (2011) seems to dispute this claim when arguing that only statistical mechanics can justify imposing a metric structure on temperature. However, Skow does not carefully distinguish between a metric structure, on which temperature *ratios* are empirically meaningful, and an interval structure, on which only ratios between temperature *differences* are empirically meaningful. It is the latter which corresponds to the group of linear transformations, and hence as far as I can see it remains possible that Nozick is right about temperature.

[8]See Pooley (2012) for a recent overview.

to sophisticated substantivalism, spacetime points have no primitive or intrinsic identities. This means that there are no two possible worlds which are qualitatively identical, but which differ over which bodies occupy which spacetime points. Since shifted models differ *only* as to which bodies are assigned which spacetime coordinates, there are no distinct possible worlds corresponding to such models. In addition, anti-haecceitism implies that there is no *cross-world identification* of spacetime points: there is no fact of the matter as to whether bodies in different possible worlds occupy the *same* spacetime point. And this means that we cannot use spacetime coordinates to consistently refer to the same spatiotemporal location across theoretical models. We can therefore naturally understand the sophisticated substantivalist as adopting QVC.

Sophisticated substantivalism is the most popular view on the nature of spacetime (Pooley 2012, 59). But the same metaphysical picture is not usually applied to properties.[9] The analogue of the sophisticated substantivalist's anti-haecceitism is *anti-quidditism* about physical magnitudes: physical magnitudes (such as being five kilogram) have no intrinsic identities, but are individuated qualitatively.[10] In order to understand how this works, it is helpful to imagine a quantity's value space as an actual space in which particular objects are located. For example, an object has a mass of five kilogram if it is located at the point in mass space which represents that magnitude. Just like physical space, mass space has a certain structure which is determined by its symmetries. If Newtonian mechanics is invariant under mass scalings, then scaling transformations are automorphisms of mass space in the same sense in which shifts are automorphisms of physical space. And this implies that we cannot obtain a distinct possible world through uniformly scaling the masses of all objects. For if there *were* such a world it would differ from the first only with respect to *which* mass properties are instantiated. But since mass properties have no quiddities, this is a distinction without a difference.

Thus, there *are* determinate mass properties on the conventionalist view,

[9] Arntzenius (2012, 58) briefly floats the idea as applied to charge before rejecting it in favour of relationism. I am unsure whether Bricker (2017) also has something like this in mind. Martens and Read (forthcoming) have also mentioned anti-quidditism in the context of (internal) symmetries.

[10] It is usual to refer to the intrinsic identities of properties (rather than individuals) as *quiddities*, but the term is used both for *quantities*, such as mass and charge, and *magnitudes*, such as five kilograms. I will use the term in the second sense. Thus, anti-quidditism about physical magnitudes means that magnitudes such as five kilograms have no intrinsic identities; but it is consistent with the idea that properties like mass do have primitive identities, such that no we couldn't swap mass with another property.

just as there are real spacetime points according to sophisticated substantivalism. But how we *refer* to these properties is conventionally determined within each model; there is no requirement that the same value refers to identical mass properties in distinct theoretical models. Indeed, if two models represent distinct possible worlds there is no fact of the matter as to whether mass values across those models represent identical mass properties.

As is clear from the above, value spaces have a rich relational structure on the conventionalist account. What grounds this structure? There are various proposals in the literature, of which I favour the introduction of second-order relations between (e.g.) mass magnitudes, such as 'x is half the mass of y'. A view like this is defended in Mundy (1987), and more recently in Eddon (2013). It may seem as if introducing second-order relations leaves no difference between the conventionalist, who believes in intrinsic properties, and the relationist, whose fundamental posits are relations such as mass ratios. But this is not true. For the second-order relations which I propose are relations between mass properties and not between the individuals which bear these properties. The relations which the conventionalist posits are thus not the relations which the relationist believes are fundamental. Indeed, it seems that most versions of relationism *also* have to impose second-order relations in order to account for facts such as that 'being twice as massive as' is half that of 'being four times as massive as'. Therefore, either view has to postulate a value space with a complex relational structure.

Finally, let me briefly mention how this picture changes once we consider gauge quantities. The conventionalist account of gauge quantities is similar to the above in that there are no cross-world comparisons between the values of gauge quantities possible. But in the case of local symmetries it is *also* impossible to compare the values of gauge quantities at different spacetime locations within the *same* world. The metaphysical implications of QVC therefore extend beyond anti-quidditism when applied to gauge theories. I will discuss these matters in detail in Chapter 6.

2.4 The Relational Structure of Our World

Relationism and conventionalism agree on which worlds are physically possible. For example, they agree we cannot generate new possible worlds through uniform mass scalings. But there also important differences between these two views. The conventionalist argues that intrinsic mass magnitudes are the fundamental properties on which mass relations supervene, whereas the relationist believes that mass relations are fundamental; if there *are* intrinsic mass magnitudes, then they supervene on mass relations. These are different ontologies, a point which Arntzenius (2012, 58) makes in the context of charge: "if charge values are fundamental then certain sets of objects are fundamental (perfectly natural), but if charge ratios are fundamental then certain sets of *ordered pairs of objects* are fundamental (perfectly natural)".

Nevertheless, there is *something* relational about conventionalism. We have already caught a glimpse of this fact: when a mass magnitude is instantiated, it is individuated through the relations it bears to other instantiated mass magnitudes. An insightful account of 'the relational structure of our world' is given in Rovelli (2014), who explores "the actual meaning of the fact that the world is described by gauge theories".[11]

Rovelli asks us to imagine a fleet of N spaceships whose dynamics depend on the ships' velocities. We can write down a set of variables x_n for their positions, but because of static shifts the *invariant* quantities are the relative positions $a_n = x_{n+1} - x_n$. It seems that if we know all the a_n, we know all there *is* to know about the fleet.

Now imagine a second fleet of spaceships with position variables y_n, and (invariant) distances $b_n = y_{n+1} - y_n$. Again, it seems that these quantities fully describe the second fleet. But if we try to describe the composite system which comprises both fleets, the sets a_n and b_n are not enough! These variables describe the distances between ships within each fleet, but not the distance between the two fleets. We thus need to introduce an additional variable, $c = y_1 - x_N$, which describes the 'coupling' between the two systems. And, tellingly, c is defined in terms of two gauge-*dependent* variables.

According to the relationist, c is a fundamental relation irreducible to the properties of either fleet considered on its own. On this view we *can* express distances in terms of position variables, but this invariably introduces redundancies in our descriptions in the form of invariance symmetries. Rovelli rejects this view. He agrees that it is *possible* to describe the composite system in terms of gauge-invariant distances, but argues that "restricting to gauge-invariant observables makes us blind to a fine structure of the world". His position is thus diametrically opposed to the relationist's insistence that invariant descriptions most faithfully capture nature's structure. An important reason for rejecting relationism is that the laws of gauge theories are expressed in terms of non-invariant quantities. For example, the semi-

[11] Rovelli's gauge symmetries are the same as what I call invariance symmetries.

classical equation of motion for a charged particle in a magnetic field involves the non-invariant vector potential field, **A**. If the fundamental quantities are invariant it is hard to account for the fact that our most successful theories have invariance symmetries.

Instead, Rovelli describes the essence of the spaceship scenario as follows:

> The position of the individual ships is redundant in the theory insofar as we measure only relative distances among these. But in the physical world each ship has a position nevertheless: this becomes meaningful with respect to one additional ship, as this appears. In other words, the existence of a gauge expresses the fact that a reference is needed to measure the position of a ship. It expresses the fact that we measure position relationally.

Rovelli states that each ship *has a position*; he thus rejects relationism. But he also argues that we cannot measure the position of a ship without relating it to other ships. In other words, we *cannot characterise position intrinsically*. There are no (non-relational) facts about *which* position a ship occupies. But this is a simple consequence of anti-haecceitism about locations! According to sophisticated substantivalism objects *have* locations, but *which* location is determined with respect to the locations of other bodies and the second-order relations between these locations. If we were to 'move' the whole universe two metres north, nothing physical would change since the relations between all instantiated location properties remain the same.

This explains the fact that our most successful theories are theories with invariance symmetries. In order to measure position we have to choose a reference, since we cannot describe locations in intrinsic terms. For example, there is no absolute centre of the universe with respect to which we can describe locations independently of how bodies occupy space. This makes it inevitable that there are countless different representations of the same physical situation, depending on our choice of reference. And this explanation also applies to non-spatiotemporal quantities. In the case of mass, for example, we conventionally choose a standard kilogram with respect to which other masses are measured.[12] Therefore, measurements of intrinsic masses are essentially relational.

[12]Note again that the point here is not that our choice of *unit* is conventional. Of course, the literalist too has to arbitrarily choose which distance counts as one metre. But she can define the metre in intrinsic terms, rather than with reference to an arbitrary standard.

It is often thought that theories with symmetries are worse than those without, all else being equal. After all, if symmetries are merely redundant structure it seems that a description of the world without symmetries better 'carves nature at its joints'. Rovelli's argument shows that, on the contrary, symmetries are a reflection of the relational nature of the world. He concludes that "gauge invariance is not just mathematical redundancy; it is an indication of the relational character of fundamental observables in physics." In other words, if quantities have no quiddities we ought to *expect* our best theories to contain invariance symmetries (as, of course, they do). Instead of aiming to reduce our theories QVC allows us to explain the importance of symmetries in physics.

2.5 Symmetries and (Superfluous) Structure

Recall that Ismael and Van Fraassen (2003) believe that symmetries encode superfluous structure: whatever structure is not invariant under the symmetries of our theories is redundant. This redundant structure includes non-invariant quantities such as spatial location and mass. But if we accept QVC, we (can) retain a realist interpretation of these quantities. What, then, is the role of symmetries according to the conventionalist?

The answer is that symmetries tell us something about the structure of the value space of particular quantities. As we saw in §2.3, physical magnitudes such as mass properties have a rich relational structure which is encoded in their second-order relations. The symmetries of our theories are a guide to what these second-order relations are. For example, the fact that mass scalings are symmetries of Newtonian mechanics tells us something about the structure of mass space; namely that mass properties stand in relations such as 'x is half of y' but not relations of the form 'x is 10 kg more than y', since the latter are not invariant under mass scalings.

There are two options for understanding this connection between the symmetries of our theories and the structure of value space. Firstly, it is possible that quantities and their relations exist independently of the laws. In this case, the laws are invariant under certain symmetries *because* the quantities which occur in these laws have a certain relational structure. But it could also be the case that quantities are individuated by the laws in which they occur. For example, mass on this view simply is whatever quantity is related to forces and accelerations as dictated by Newton's Laws (Dees (2018) defends a view like this). On this option the grounding relation goes in the opposite direction: quantities have a structure which reflects the symmetries of the laws which govern them, rather than vice versa.

I am neutral between these two options. In either case, symmetries tell us something positive and interesting about the structure of those quantities which feature in our theories, rather than something essentially negative about which structures are superfluous. If all symmetries do is encode excess structure then their importance to modern physics is a puzzling fact, as Rovelli remarks. But if symmetries are, in a sense, part of the structure of nature, this puzzle disappears. On the conventionalist approach, symmetries are our guide to the structure of physical quantities.

Chapter 3

Against Invariance

It is time to return to the Inconsistent Triad. Recall that (2) of the Inconsistent Triad states that gauge quantities are never physically real. In Chapter 1, I argued that (2) was motivated by the *Invariance Principle*: only invariant quantities are physically real.

In this chapter I argue against the Invariance Principle, and hence against (2). As I will show below, the Invariance Principle is inconsistent with Quantity-Value Conventionalism. Since the latter position has no presumption against gauge quantities, it allows us to decide whether such quantities are physically real on the basis of their explanatory value. In the remaining chapters, I will argue that gauge quantities *are* explanatory, and hence physically real.

The plan for this chapter is as follows. Firstly, I distinguish between the Invariance Principle and what is called *Leibniz Equivalence* (§3.1). These two principles are often conflated, but I will show that the Invariance Principle follows from Leibniz Equivalence only if we assume Quantity-Value Literalism (QVL). I will then discuss three prominent arguments for the Invariance Principle (§§3.2-3.4): the *metaphysical* claim that interpreting gauge quantities as physically real leads to radical indeterminism (Earman and Norton 1987; Wallace 2002); the *epistemic* argument that gauge quantities are in principle undetectable (Dasgupta 2016); and lastly the *semantic* claim that gauge quantities are multiply realised, and therefore do not (uniquely) refer (Healey 2006; Caulton 2015).[1]

I argue that each of these arguments at most establishes Leibniz Equiva-

[1] In fact, the latter two arguments apply to *any* non-invariant quantity. I will therefore also speak more generally of non-invariant quantities, which includes the relevant gauge quantities.

lence. But Leibniz Equivalence is compatible with Quantity-Value Conventionalism (QVL), and hence with a realist interpretation of gauge quantities. Therefore, these arguments do not force us into relationism.

3.1 Leibniz Equivalence

We have already encountered the claim that symmetry-related models are physically equivalent, for example in the case of static shifts or mass scalings. Recall also Ismael and Van Fraassen's (2003) idea of symmetries as a guide to superfluous structure. We can state a general version of this claim as follows:

> *Leibniz Equivalence*: If two theoretical models are related by a symmetry transformation, then they are physically equivalent.[2]

The exact content of Leibniz Equivalence depends on which symmetries we intend it to cover. We can either fill this in on a case-by-case basis or offer a general characterisation of invariance symmetries. In this thesis, I will opt for the former approach and only consider specific examples such as static shifts and mass scalings. Nevertheless, the argument I give holds for all symmetries satisfying Leibniz Equivalence.

Leibniz Equivalence and the Invariance Principle are closely related: as we have seen in the previous chapter, the physical equivalence of symmetry-related models often motivates us to get rid of non-invariant quantities, such as absolute position. Nevertheless, the two principles *are* distinct, a fact which has often gone unnoticed.[3] For example, Saunders (2007, 453) equivocates between them in the following passage:

> [Relationalism] says that *only quantities invariant under exact symmetries are real*—thus relative directions, relative distances, and so on, under rotations and translations, etc. [...]. Call [this principle] the invariance principle. The distinctions among representations then correspond to nothing physically real. *Equivalently, such models, as goes their physical content, can be simply identified.* (emphasis mine)

[2]Cf. Earman and Norton (1987, 522).

[3]An exception is Møller-Nielsen and Read (forthcoming, §6), who point out the inequivalence of the two principles.

The first italicised part of this passage is a canonical statement of the Invariance Principle. But at the end of the passage, Saunders equates this principle with Leibniz Equivalence, the claim that symmetry-related models are identical in terms of their physical content. However, Saunders does not offer an argument for this supposed equivalence, and neither am I aware of any other philosopher giving such an argument.

This is significant, because the arguments for the Invariance Principle which I discuss below all seem to have the following form: (a) to avoid indeterminism/undetectability/non-reference in our theories, we need to accept Leibniz Equivalence; (b) Leibniz Equivalence implies the Invariance Principle; therefore, (c) to avoid indeterminism/undetectability/non-reference, we need to accept the Invariance Principle. If (b) is false and Leibniz Equivalence does not imply the Invariance Principle, then these arguments fail to justify the latter.[4]

I believe that Leibniz Equivalence only implies the Invariance Principle if, in addition, we assume that QVL is true. If we accept QVC, on the other hand, we can consistently accept Leibniz Equivalence while rejecting the Invariance Principle. Therefore, if I am right that Leibniz Equivalence is enough to avoid the spectres of indeterminism, undetectability and non-reference, it follows we can do so while maintaining a realist interpretation of non-invariant quantities.

As far as I am aware there are no explicit arguments in the literature for the claim that Leibniz Equivalence implies the Invariance Principle, but I believe that the following line of reasoning is often implicitly assumed. Let $\mathcal{M}$ and $\mathcal{M}'$ be symmetry-related models of a theory T, such that there is a non-invariant quantity Q which takes on different values in $\mathcal{M}$ and $\mathcal{M}'$. Then:

(i) $\mathcal{M}$ and $\mathcal{M}'$ represent the same physical possibility (Leibniz Equivalence);

(ii) If two models represent the same physical possibility, they agree on all physical features;

(iii) $\mathcal{M}$ and $\mathcal{M}'$ disagree on the value of Q;

(iv) Therefore, the values of Q cannot represent the same physical features in $\mathcal{M}$ and $\mathcal{M}'$;

[4] The Invariance Principle *does* imply Leibniz Equivalence: the non-invariant quantities are exactly what varies across symmetry-related models, so if these are not real then symmetry-related models must have the same physical content.

QVL Quantities are physically real only if their values consistently refer to particular physical magnitudes across theoretical models;

(C) Therefore, Q is not a physically real quantity.

Since we have not assumed anything about Q, the conclusion holds for arbitrary non-invariant quantities, and thus is equivalent to the Invariance Principle.

It is obvious that this argument is not valid without assuming QVL, so Leibniz Equivalence and QVL together imply the Invariance Principle. Furthermore, Leibniz Equivalence does not imply the Invariance Principle if we *don't* assume QVL. I will prove this by showing that Leibniz Equivalence and QVC (which is inconsistent with QVL) implies the *negation* of the Invariance Principle.

Suppose that the conventionalist accepts Leibniz Equivalence for some symmetries. For example, suppose she believes that shift-related models are physically equivalent. This is consistent with QVC, since the Meaningfulness Criterion only commits the conventionalist to those statements which are invariant across symmetries. But it then follows from the Meaningfulness Criterion that there are physically meaningful statements about non-invariant quantities. One such statement is that every body has a spatial location, and this means that spatial location is a physically real quantity. Since location values are non-invariant under static shifts, this realist commitment implies the negation of the Invariance Principle. Therefore, Leibniz Equivalence, QVC and the Invariance Principle are inconsistent.

I will now argue that to avoid the spectres of indeterminism, undetectability and non-reference we at most need to accept Leibniz Equivalence, and hence these arguments do not justify the Invariance Principle.

3.2 Invariance and QVL (I): Determinism

The claim that gauge symmetries lead to indeterminism in our theories is well-known. Earman and Norton (1987) argue that this happens in General Relativity due to the Hole Argument, and Belot (1998) applies a similar line of reasoning to electromagnetism. Wallace (2002) generalises the argument to any (time-dependent) gauge symmetry. The thrust of these arguments is that there are distinct models which agree on the physical facts at all times before t, but which differ by a gauge transformation after t. If we believe that these models represent distinct physical possibilities, then this implies a breakdown of determinism: the history of the universe up to time t does not determine its states at later times.

Earman and Norton (1987) argue that we should avoid this sort of radical indeterminism not because indeterminism is in principle unacceptable, but because "[determinism] should fail for a reason of physics, not because of a commitment to substantival properties which can be eradicated without affecting the empirical consequences of the theory" (524). In a spacetime context, the 'substantival properties' to which Earman and Norton refer here are spacetime points. Their conclusion is that we have to accept Leibniz Equivalence for spacetime symmetries, and that therefore the fundamental spatiotemporal quantities are distances between bodies rather than intrinsic locations.

I agree with Earman and Norton that accepting Leibniz Equivalence is necessary if we want to avoid this sort of indeterminism.[5] However, it does not follow that we have to 'eradicate' substantival properties, or more generally non-invariant quantities, from our theories. This is only a valid inference if we accept QVL. If we accept QVC, on the other hand, Leibniz Equivalence implies that we can remain realist about non-invariant quantities. In the case of spacetime, for example, sophisticated substantivalists accept that symmetry-related models are physically equivalent, but are nevertheless realist about spacetime. I think the same strategy can be used in the case of internal gauge symmetries, such as the gauge freedom of electrodynamics. Here too we can simply understand gauge-related models as distinct representations of the *same* physical possibility, rather than as representations of different possible futures given a fixed past. We can explicate this in terms of an anti-haecceitist oranti-quidditist ontology, as I will do in Chapter 6. This suffices to avoid indeterminism, since the gauge-related models which were supposed to represent distinct possible futures given a determinate past are now considered as distinct representations of the same *unique* future of that past. Therefore, we need not accept the Invariance Principle in order to avoid radical indeterminism if we subscribe to Quantity-Value Conventionalism.

[5] But I am unsure why indeterminism could not be the result of an interpretative choice; to me, it seems that interpretations *should* have interesting consequences about what the world is like. For example, it seems odd to claim that we should reject Bohmian mechanics because whether quantum mechanics is deterministic ought not to depend on a metaphysical commitment to corpuscular particles "which can be eradicated without affecting the empirical consequences of the theory".

3.3 Invariance and QVL (II): Detectability

The second argument for the Invariance Principle I will discuss is based on *epistemic* grounds. According to Dasgupta (2016), we should renounce non-invariant quantities because they are in principle *undetectable*. On his account, the *symmetry-to-reality inference* goes as follows (Dasgupta 2016, 843):

(i) Laws L are the complete laws of motion governing our world.

(ii) Feature X is variant under the symmetries of L.

(iii) Therefore, X is undetectable (from (i) and (ii)).

(C) Therefore, X is not real (from (3) and an Occamist norm that we dispense with undetectable structure).

Again, I will argue that we can avoid the undetectability of non-invariant quantities by assuming Leibniz Equivalence; but this only implies the Invariance Principle if we also assume Quantity-Value Literalism. Firstly, however, let me remark on the 'Occamist norm' to which Dasgupta defers. The norm states that *all else equal* we should dispense with undetectable structure. The 'all else equal' clause is important here: Dasgupta believes that there may be good reasons to posit undetectable structure. For example, such structure may yield a simpler theory, or one with greater explanatory strength. "In that case, we would have empirical evidence of sorts that the feature is real, in the sense that our all-things-considered best empirically confirmed theory implies that it is real" (Dasgupta 2016, 854). Dasgupta thus believes it is possible that gauge quantities are explanatory, and that if so we have good reason to believe in their physical reality even if they are undetectable. I will argue in the following chapters that some non-invariant quantities *are* explanatory; hence, Dasgupta's undetectability argument fails to make a case against (3) if we follow it to the letter. However, the spirit of the argument is very much against non-invariant quantities and we can easily imagine a stronger position which holds that there are *no* good reasons to posit undetectable quantities. It is therefore worthwhile to show that on the conventionalist view it is not the case that variant quantities are undetectable.

Firstly, here is how Dasgupta initially defines undetectability:

> [S]omething is undetectable in my sense if, roughly speaking, it

> is physically impossible for it to have an impact on our senses. (Dasgupta 2016, 854)

This definition of detectability is rather weak. For example, even absolute velocities are detectable on this definition: after all, in Newtonian mechanics relative velocities supervene on absolute velocities, and since the former impact our senses, so do the latter. However, Dasgupta quickly revises this initial definition in favour of something stronger:

> It then follows [...] that the feature is undetectable [if] there is no physically possible process by which we might discover which determinate values are actually instantiated.

This is a stronger criterion, since it demands that we can measure the particular *value* of a quantity (presumably up to some degree of accuracy). Let us call this notion *measurability*. As the example of absolute velocity shows, it is possible for something to have an impact on our senses without being measurable. Thus, what is undetectable according to Dasgupta are not the non-invariant quantities themselves, but their particular values. Suppose, for example, that the current absolute velocity of the Taj Mahal is 100 km/h. And suppose that there is a possible world which is qualitatively identical to ours, but in which the whole universe has been boosted such that the velocity of the Taj Mahal is now 200 km/h. Since these two worlds are empirically the same, there is no experiment which could reveal which absolute velocity the Taj Mahal has in the actual world. For such an experiment, if it were an accurate measurement of the Taj Mahal's absolute velocity, would have different observable outcomes in these two worlds. Therefore, what we cannot know even in principle is whether the Taj Mahal has the absolute velocity denoted by '100 km/h', rather than the absolute velocity denoted by '200 km/h', and *mutatis mutandis* for other non-invariant quantities.

If we follow Dasgupta's Occamist advice and reject the physical reality of absolute velocity, then there are no absolute velocity values to measure, and hence no unmeasurable quantities to worry about. However, this solution to the problem of undetectability is unnecessarily strong. For as we have seen, the problem arises when our theories posit worlds which are empirically equivalent yet physically distinct. The problem is a failure of Leibniz Equivalence. If symmetry-related models represent the same possible worlds, then there are no unmeasurable differences between them. For example, if boosted models are physically equivalent, then there is no meaningful question as to whether the Taj Mahal is moving at 100 or 200 km/h (or better, the velocities which we refer to in our world as '100 km/h' and '200 km/h').

According to Dasgtupa's argument it is undetectable *which* absolute velocity an object has. But if this is a question about the intrinsic identities of velocity properties, then it simply has no answer on a conventionalist account of quantities. The labels '100 km/h' and '200 km/h' in boost-related models are exactly that: different names for the same velocity, which is qualitatively individuated. Of course, we can measure the *relative* velocity of an object, and these relative velocities together determine the identities of all intrinsic velocity properties. In this sense we *can* detect velocities; but it is conventional which names we give them.

Again, there are analogous arguments for non-spatiotemporal properties such as mass. We can't measure intrinsic mass identities, for there are none, and it is meaningless to ask whether we live in the world in which the dome of the Taj Mahal has the mass which we refer to as 12 500 tonnes or in a qualitatively identical world in which it weighs double that, for these worlds are the same. But we *can* measure intrinsic masses through measuring the ratios which determine their identities, which is in effect what we do when we measure the mass of an object in a relational unit, such as kilograms, which is defined with reference to a standard mass.

Therefore, the conventionalist can avoid undetectable properties without accepting the Invariance Principle, simply by positing Leibniz Equivalence. And hence Dasgupta's argument from undetectability does not offer a good justification for the claim that only invariant quantities are physically real.

3.4 Invariance and QVL (III): Reference

The previous section was concerned with an epistemic argument. The next argument, found in Healey (2006) and Caulton (2015), is *semantic* in origin. The claim I will contest is that non-invariant quantities have no unique referent, and therefore cannot be given a realist interpretation.

The argument goes as follows. Suppose we are trying to interpret a physical theory. At least part of the task of interpretation is to assign referents to the terms which occur in that theory. Some of these quantities will obviously correspond to physical quantities with which we are directly acquainted, such as colour and distance. We call these entities *observables*. Since we already know what the observables *are*, we can simply stipulate which quantities refer to them. However, theories usually also introduce new terms for unobservable quantities, and we have no grasp on what these unobservables are apart from what the theory says about them. It seems, then, that we cannot know what these *theoretical* terms, such as 'electric field', refer to, unless we have, *per impossibile*, *already* interpreted the theories in which they occur.

Fortunately, the situation is not hopeless. Building on the work of Lewis (1970), we can construct implicit definitions of theoretical terms once we have interpreted the observable terms by stipulation. Instead of directly connecting a theoretical term to its referent, we declare that the term refers to *whatever* it is that is related to the observables in the manner described by our theories. For example, the electric field is just *whatever* is generated by, and exerts a force on, charged particles in accordance with Maxwell's Laws. If there are multiple uninterpreted theoretical terms, then the implicit definition says these refer to those elements of nature (whatever they are) which are related to the observables *and* to each other in accordance with the laws. If we have a syntactic formulation of our theory, then this gives us none other than the Ramsey sentence, which is the result of replacing theoretical terms with variables which are then existentially quantified over.

For this approach to work, our theories must be *uniquely realised*. That is, there must exist a unique set of unobservable quantities which the theoretical terms of our theories refer to. If this condition does not hold, then either the theory is not realised at all, i.e. it is false, or it is *multiply realised*. In the latter case it is indeterminate what theoretical terms such as 'electric field' refer to, and hence it is impossible for us to know what an electric field is, given that our only grasp on such terms comes from the theories in which they occur. Healey and Caulton argue that if we interpret non-invariant quantities as physically real, theories with non-trivial symmetries are multiply realised. Therefore, if we want to avoid indeterminacy of reference we have to embrace the Invariance Principle. As with the previous two arguments, I argue that accepting Leibniz Equivalence is enough to avoid non-reference; the step from Leibniz Equivalence to the Invariance Principle is unwarranted.

I will borrow an example from Healey (2006) to illustrate this. Suppose we have a toy theory in which sub-atomic particles, quarks, cluster together to form protons and neutrons. Quarks can have one of three colours: red, green and blue. Colour is a dynamically efficacious property which couples to various other quantities according to the laws. However, the theory has a colour permutation symmetry, such that if we change (for example) all red quarks to green, all green quarks to blue, and all blue quarks to red, we obtain an empirically indistinguishable model. Finally, the dynamics of our toy theory are such that quarks are strongly confined in colour-neutral combinations of red, green and blue, and thus it is impossible to observe single quarks on their own, or groups of quarks which do not have an even distribution of colours.

Healey argues that in this set-up, the terms 'green', 'red' and 'blue' are referentially indeterminate. Suppose that there are three properties, which we denote R, G and B, and that our toy theory has a partially interpreted model of the form $\mathcal{M}(r, g, b)$, where r, g and b are variables standing for the colour properties of quarks. We now want to implicitly define r as whatever property satisfies $\mathcal{M}$, and *mutatis mutandis* for g and b. Suppose that $\mathcal{M}$ is satisfied if r refers to R, g refers to G, and b refers to B. In other words, $\mathcal{M}(R, G, B)$ is a dynamically possible model. It then seems that the implicit definition of r picks out R as its unique referent, and similarly for the other terms. However, since our theory is colour permutation invariant $\mathcal{M}(G, B, R)$ is *also* a dynamically possible model. And this means that $\mathcal{M}(r, g, b)$ is *also* satisfied if r refers to G, and g refers to B, and b refers to R. Therefore, an implicit definition in terms of $\mathcal{M}(r, g, b)$ leaves it indeterminate whether r refers to R or G (or even B!), and similarly for g and b. This is a direct consequence of the fact that $\mathcal{M}(r, g, b)$ is multiply realised, which is the result of our theory's colour permutation symmetry.

Healey and Caulton conclude from this that non-invariant quantities such as colour are not physically real. Caulton (2015, 161), for example, writes that this "prompts appropriate reform towards a *new* formalism, in which the physical properties and relations—including the unobservable ones—are transparently represented without redundancy". In effect, Caulton proposes that we construct reduced theories in which no non-invariant quantities occur. However, this demand is too strong: assuming Leibniz Equivalence is enough to avoid non-reference. And, as we now know, the Invariance Principle only follows from Leibniz Equivalence on the assumption of QVL. I will now show how we can avoid non-reference on a conventionalist account.

The problem which causes the indeterminate reference of the terms r, g and b is that $\mathcal{M}(r, g, b)$ is multiply realised: the symmetry-related models $\mathcal{M}(R, G, B)$, $\mathcal{M}(G, B, R)$ and $\mathcal{M}(B, R, G)$ are its distinct realisations. But if we accept Leibniz Equivalence symmetry-related models are distinct representations of the *same* state of affairs, and hence the difference between $\mathcal{M}(R, G, B)$, $\mathcal{M}(G, B, R)$ and $\mathcal{M}(B, R, G)$ is merely representational. But if this is true then $\mathcal{M}(r, g, b)$ is *not* multiply realised: it only has one physical realisation, which is represented by a class of symmetry-related models.

This means that if we accept Leibniz Equivalence, we can avoid the multiple realisation of which the indeterminate reference of theoretical terms is a consequence.

On this picture, what properties *do* the terms r, b and g refer to? Here we face a similar dilemma as in the previous section when discussing detectability. If the question asks after the intrinsic identities of the referents of these theoretical terms, it is meaningless: there is no colour property to which we can consistently refer across theoretical models with r, b or g. Whether we call a colour property 'red', 'green' or 'blue' is a conventional choice which may vary across models. Nevertheless, it *is* possible to identify theoretical terms qualitatively. For example, we can (arbitrarily) stipulate that r refers to the colour of the quark which is located at the centre of mass of a proton. This is analogous to stipulating that we will use the term 'kilogram' to refer to the mass of the standard kilogram in Paris.

Healey argues that even reference by ostentation will fail if the constitution of protons is highly symmetric. For example, if all three quarks in a proton are co-located at one point and share all their other qualitative properties, then we cannot give any description which picks out a unique colour property. But note that this is a highly unusual situation; it is analogous to an empty spacetime in which all points are qualitatively identical, or to Max Black's famous example of two identical smooth black spheres located some distance apart from each other. In these cases, there is no description which picks out one of the two spheres, or one point of spacetime over another. However, we *can* give descriptions which pick out both objects at once. For example, we can describe Black's scenario as follows: there are two smooth black spheres which are some distance apart from each other. This description *is* uniquely realised. Therefore, we can overcome the threat of multiple realisations even in cases of qualitatively identical objects.

3.5 Conclusion

I have analysed three arguments which are offered as justifications of the Invariance Principle: that we have to accept the Invariance Principle in order to avoid indeterminism, undetectability and non-reference, respectively. However, as I have shown in §3.2-3.5, we at most need to assume Leibniz Equivalence in order to avoid these spectres. Leibniz Equivalence and the Invariance Principle are closely related principles but they are not equivalent (§3.1). If we adopt Quantity-Value Conventionalism instead of Quantity-Value Literalism, we can consistently accept Leibniz Equivalence and reject the Invariance Principle. This allows us to interpret non-invariant quantities as physically real while avoiding indeterminism, undetectability and non-reference.

The orthodoxy of the Invariance Principle thus seems to rest on the mistaken belief that it is equivalent to Leibniz Equivalence. For the purposes of this thesis, this conclusion is important mainly because it allows us to directly evaluate whether gauge quantities are *explanatory*. If there *are* explanatory gauge quantities, then they are physically real according to (1) of the Inconsistent Triad.

In the remainder of this thesis I will show that interpreting gauge quantities as physically real *does* give our theories greater explanatory strength. In the next chapter, I will explain in more detail what sort of explanatory strength this is. In Chapter 5 and 6, I apply the foregoing considerations to the Aharonov-Bohm effect, arguing that we should interpret the vector potential field as physically real.

Chapter 4

Cosmic Coincidences and Local Action

In the previous two chapters, I have presented an alternative position to the prevailing Quantity-Value Literalism, which I have called Quantity-Value Conventionalism. I have shown how this approach allows us to be realists about non-invariant quantities while avoiding the three-fold spectres of indeterminism, undetectability and non-reference.

However, I have not yet given any positive reason for adopting conventionalism. This is what I plan to do in this chapter. The standard alternative to conventionalism is relationism. For example, faced with the invariance of Newtonian mechanics under static shifts the literalist must adopt an ontology of fundamental spatial distances in order to avoid the spectres mentioned above; similarly, the literalist posits fundamental mass relations in response to invariances under mass scalings. In this chapter, I will give an argument for preferring non-invariant intrinsic properties over invariant fundamental relations.

The argument is that non-invariant quantities can explain relations between objects which otherwise seem like 'cosmic coincidences'. I will focus on the case of mass: as I will argue, intrinsic mass properties can explain the so-called 'conspiracy of mass relations'. However, I believe that the argument generalises to some other non-invariant quantities, and I will argue in Chapter 6 that the gauge dependent vector potential can similarly explain relations between loop holonomies which are otherwise posited as brute facts.

The structure for this chapter is as follows. In §4.1, I return to the No Miracles Argument for scientific realism and analyse when miracles demand explanations. In §4.2, I explain the conspiracy of mass ratios. Then, in §4.3, I pose the following dilemma to the comparativist about mass: either the conspiracy of mass ratios is a cosmic coincidence which we ought to explain by positing fundamental mass properties, or mass relations act on each other at a distance and so violate the principle of local action. Finally, in §4.4 I spell out how the conventionalist explains the conspiracy of mass ratios.

4.1 Of Cosmic Coincidences

The intuition behind the No Miracles Argument for scientific realism is that if our theories were false, the fact that they are so empirically successful would be a miracle. Therefore, our theories are true.[1] In a similar vein, J. J. C. Smart talks of a *cosmic coincidence*:

> Is it not odd that the phenomena of the world should be such as to make a purely instrumental theory true? On the other hand, if we interpret a theory in a realist way, then we have no need for such a cosmic coincidence: it is not surprising that galvanometers and cloud chambers behave in the sort of way they do, for if there really are electrons, etc., this is just what we should expect. (Smart 1963, 39)

Hacking (1981) gives another variant of this argument, which he calls the 'Argument of the Grid'. He describes a grid which is drawn on normal-sized paper, with numbers written in the corners. This grid is then reduced photographically, and microscopic versions of it are prepared. The lines and numbers on these micro-grids are not visible to the human eye, but when viewed through a microscope the numbers can be clearly seen in the corners of the grid. I suspect anyone reading this argument finds it extremely compelling to believe that there is in fact a microscopic grid under the microscope. As Hacking puts it:

> Can we entertain the possibility that, all the same, this is some gigantic coincidence? Is it false that the disc is, in fine, in the shape of a labelled grid? Is it a gigantic conspiracy of 13 totally

[1] On a more sophisticated version, the argument is that the success of science itself is a miracle, which is explained by scientific realism in general.

> unrelated physical processes that the large scale grid was shrunk into some non-grid which when viewed using 12 different kinds of microscopes still looks like a grid? To be an anti-realist about that grid you would have to invoke a malign Cartesian demon of the microscope.

In this section, I will make this version of the No Miracles Argument a little more precise. We can extract the following two key features: firstly, that things look *as if* our theories are true, and secondly that if they were *not* true, things *could easily have been different*. If the phenomena could have been different, but just happen to make it look like our theories are correct about the unobservable, then this is a seeming coincidence which realism can explain.

The idea that the phenomena *look like* the theories describing them are true is explicit in both passages cited above. Smart argues that it looks as if there are electrons since if there *were* electrons, we would expect to observe exactly the phenomena we in fact do observe. Similarly, it looks as if there is a microscopic grid under Hacking's microscope because if there were such a grid, we would expect the microscope to produce grid-like images. We can put this more technically as follows: the phenomena are as if our theories are true since models of the empirical phenomena are (isomorphically) embeddable into models of our theories (cf. Van Fraassen (2008)).

The observation that the world looks like our theories are (approximately) true is not enough to sustain the No Miracles Argument. For example, our empirical observations are *as if* we are living in a simulation, in the sense that if we *were* living in a simulation, we would not expect our observations to differ (Bostrom 2003). But this hardly gives us reason to believe that we are in fact living in a simulation, so something else is needed for the No Miracles Argument to work.

The second step of the argument, then, is that if our theories were false, the observable phenomena easily could have been different. This claim is explicit in the argument of the grid. According to Hacking, it is "a gigantic conspiracy" if different microscopes employing unrelated physical processes reliably produce grid-like images without there being an actual microscopic grid. If these processes are independent, each one *could have* produced a different image. The fact that all these processes seem to 'coordinate' makes it look like the microscopes conspire together. On the other hand, if we assume that there *is* a grid, the microscopes don't have to conspire: each of them is simply producing an image of what is in front of them. What is at work here is something like a common cause principle: correlations between distant events require a common cause.

But we soon run into difficulties when spelling out this second step. For an anti-realist can easily take any theory, T, and propose a new theory T' which states that T is empirically adequate. If T' is true, then the observable facts could *not* have been different, at least not unless T' (and hence T) were false. For example, if T' is the claim that our theories of microscopes and grids are empirically adequate, then the twelve microscopes in Hacking's example could not have produced anything other than grid-like images, since this is what T' predicts. If this gambit works, the No Miracles Argument loses all its force — so the realist has to have a response.

I will not venture a general solution to this puzzle, as it is not my goal here to defend scientific realism. Instead, I will focus on a particular example of what looks like a cosmic coincidence, namely what Martens (2019a) calls the *conspiracy of mass relations*. As we will see below, this case is both more tractable and highly relevant to the debate between QVL and QVC. I will argue that the literalist who believes in fundamental mass relations either has to postulate a cosmic coincidence, or else violate a principle of local action.

4.2 The Conspiracy of Mass Relations

According to the comparativist about mass, mass relations are fundamental; if there *are* intrinsic mass properties, then these supervene on mass relations. As we saw in Chapter 2, Quantity-Value Literalism leads to comparativism about mass once we identify models related by mass scalings. The value space of these mass relations is (isomorphic to) the positive real numbers, $\mathbb{R}^+$, and in principle any pair of objects could stand in any of these mass relation to each other.

In fact, however, mass relations seem to obey the following constraint:

$$m_{ik} = m_{ij} \cdot m_{jk}, \tag{4.1}$$

where m_{ij} is the mass relation between two objects i and j. I will call (4.1) the *transitivity of mass relations* (TMR).[2] I claim that TMR is the sort of cosmic coincidence which demands explanation. Firstly, as I will argue below, TMR makes it seem *as if* the mass relations are mass *ratios* which supervene on intrinsic masses. Secondly, as I will argue in the next

[2]Martens (2019a) calls this the *conspiracy of mass relations*. Martens and Read (forthcoming, §5.2) also discuss how intrinsic masses can explain that (4.1) holds.

section, TMR easily could have failed to hold *unless* we posit the existence of intrinsic masses. Therefore, it follows from the No Miracles Argument that there *are* intrinsic masses.

If there are intrinsic masses, TMR becomes a mathematical theorem. For if mass relations are ratios between mass properties, we can write $m_{ij} = m_i/m_j$, and hence:

$$m_{ik} = \frac{m_i}{m_k} = \frac{m_i}{m_j} \cdot \frac{m_j}{m_k} = m_{ij} \cdot m_{jk} \tag{4.2}$$

It is clear that this equality holds whatever values we assign to m_i, m_j and m_k. In other words, TMR is a necessary condition for representing mass relations with intrinsic mass values; if TMR fails, there is no way of assigning mass values to objects which faithfully captures all mass relations. Hence, we can only embed comparativist models of mass relations into models of absolute mass values if TMR holds. Therefore, the fact that TMR *does* hold makes it seem *as if* there are intrinsic masses.[3]

This is the first step of the No Miracles Argument for intrinsic masses. The second step is this: it would seem like a cosmic coincidence that the mass relations between distinct pairs of objects are transitive, given that in principle they could have taken on any value from $\mathbb{R}^+$. For example, m_{ij} could have been slightly higher, in which case TMR would have failed to hold, all else remaining equal. Surely, distant objects do not *conspire* with each other to have their mass relations look just like ratios of intrinsic masses? If, on the other hand, we assume that there *are* intrinsic masses, the transitivity of mass ratios is not a conspiracy at all, but a straightforward mathematical fact which could not have failed to hold. Therefore, we ought to avoid postulating a cosmic coincidence and instead posit intrinsic masses on which mass ratios supervene.

But the comparativist can push back and argue that m_{ij} could *not* have had a different value without the values of other mass relations also changing. She can do this by proclaiming the transitivity of mass ratios a law of nature, which is not so easily violated.[4] If TMR is a law, then its holding is no more a cosmic coincidence than the fact that all massive objects happen to accelerate when acted on by a force. Of course, this is the same issue we encountered in our earlier discussion of the grid argument. I will now argue

[3]This is not a novel argument; the same point is made in Martens and Read (forthcoming).

[4]Dees (2015, 129) suggests that the relationist can explain the triangle inequality, which is the analogue TMR but for distances, with a similar move: "The distance-firster could hold that the triangle inequality holds only with *nomological* necessity".

that the comparativist's response commits her to an endorsement of action at a distance, and therefore does not succeed as a defence of fundamental mass relations.

4.3 Local Action

I will argue that the comparativist faces the following dilemma: either it could easily have been the case that mass relations are not transitive, or mass relations causally act on each other at a distance. If the former is true then TMR is a cosmic coincidence, and we need to postulate intrinsic masses to explain it. If the latter is true, then comparativism violates the principle of local action, which is one of the most firmly held metaphysical principles in physics. I doubt that any comparativist is willing to pay this price.

Let's start with the first horn of this dilemma. Consider the counterfactual situation in which the mass relation between a and b is slightly lower than it is now. This possibility easily could have obtained: it is both physically possible for m_{ab} to have a slightly different value, and a situation not too far removed from the actual world. What would have happened to the other mass relations in this case? If they remain the same, then TMR is immediately violated. After all, in the actual world $m_{ab} = m_{ac} \cdot m_{cb}$. If we change the value of m_{ab} without changing the values of m_{ac} or m_{cb}, then this equality no longer holds. In this case, then, TMR *could* easily have failed, and the fact that it holds in the actual world is a miraculous coincidence.

There are good reasons to believe that if m_{ab} had been different, m_{ac} and m_{cb} would have retained the same value. Lewis (1979, 472) argues that in analysing counterfactuals, our first priority is to "avoid big, widespread, diverse violations of law", but our second priority is to "maximize the spatiotemporal region throughout which perfect match of particular fact prevails". And only *then* is it important to avoid "small, localized, simple violations of law". Now, changing only the value of m_{ab} leaves facts about the large spatiotemporal region which excludes a and b invariant. On the other hand, if we also change other mass relations, we change particular facts throughout spacetime. Thus, upholding TMR violates Lewis' second condition. The question, therefore, is whether a violation of TMR is "big, widespread [and] diverse" or "small, localized [and] simple". I argue it is the latter, since it is only the mass relation between a and b which is 'out of step'; all other mass relations remain transitive with each other. The violation thus seems small and simple, contained to a singular incident, rather than big and widespread. We therefore have a trade-off between maximising a perfect match of particular facts and small violations of the laws. If Lewis is right, we should prefer the latter. Therefore, if m_{ab} had been slightly different — a situation which easily could have obtained — TMR would have failed, and hence the fact that it holds in the actual world is still a miracle.

But suppose the comparativist insists that TMR cannot easily fail: had m_{ab} been different, other mass relations too would have changed. If this were the case, mass relations would causally act on each other at a distance. We can leverage Lewis' (1973) analysis of causation in terms of counterfactuals to spell out the relevant sense of causal dependence. According to Lewis, an event c is causally dependent on an event e iff either (a) c and e actually occur and had c not occurred, then e would not have occurred, or (b) c and e don't actually occur, but had c occurred, then e would have occurred.[5]

Let c stand for a and b's having the mass relation m_{ab}, and e stand for a and c's having m_{ac}. Then, according to the comparativist, were c not to obtain — that is, were m_{ab} slightly lower — e would not have obtained, since m_{ac} needs to change value for TMR to hold. But, on Lewis' analysis, this means that m_{ab}'s value *causes* m_{ac}'s value. There is a necessary connection between m_{ab} and m_{ac}, in that if we were to change the former the latter would also change. And this is plainly action at a distance, since m_{ab} is an irreducible fact about a and b, and m_{ac} is an irreducible fact about a and c. But the region occupied by a and c and the region occupied by a and c don't fully overlap, and so a mass relation in one region causes a change somewhere where it is not.

The comparativist thus faces the following dilemma: either the transitivity of mass relations easily could have failed, even if it were a law, or it is a *causal* law which implies that mass relations act on each other at a distance. I take it that no comparativist wants to defend action at a distance; the principle that causes always act locally is one of the oldest and most widely-held principles of physics (Hesse 1955). This leaves the comparativist the other horn of the dilemma, on which TMR is a cosmic coincidence which makes it look as if there are intrinsic masses. Therefore, there is a clear explanatory benefit to positing intrinsic masses.

[5]The sprawling literature on causation contains many counterexamples to and further refinements of Lewis' approach, as well as a wild variety of alternative approaches, but I lack the space to even scratch the surface of these debates. Similarly, I won't go into much detail about how to analyse the counterfactuals which figure in Lewis' definition of causation.

4.4 The Conventionalist Explanation

It seems easy to understand how a literalist about mass properties can explain TMR. For the literalist, each mass magnitude is associated with a unique dimensionful number, such as 5 kg or 5000g. The transitivity of mass ratios then follows from the relations between these numbers. But the conventionalist seems to face the following problem. According to the conventionalist, values such as 5 kg or 5000g do *not* uniquely refer to particular mass magnitudes — it is conventional whether we use them to refer to this or that mass. But if masses don't come 'attached' with (dimensionful) numbers, we can't appeal to those numbers to explain TMR.

However, the above reasoning misrepresents the literalist. Although the literalist believes that mass values refer to the same magnitudes across theoretical models, she is not committed to *identifying* mass magnitudes with numbers or relations to numbers. As Martens (2019b) remarks, there is nothing inherently five-ish about a five-kilogram object, unlike the way in which a five-fingered glove exemplifies the number five. Since TMR is a relation among numerical values, even the literalist has to explain something about how we come to represent mass magnitudes numerically. Furthermore, while it is true that according to the literalist the same value uniquely refers to a particular mass magnitude, it is *not* the case that a magnitude is denoted by a unique dimensionful *number*. For example, we can denote the mass of a five-kilogram object with 5 kg, but also with 5000g or 11 lbs. These are different representations, not different properties. Therefore, we cannot straightforwardly appeal to the nature of intrinsic mass properties to explain numerical relations between them.

On the view which I prefer (see §2.3), there are second-order relations between mass properties, such as one mass being twice that of another.[6] We can then prove representation theorems to the effect that these relations are faithfully represented by various mass scales (Ellis 1966; Mundy 1987); for example, the second-order relation 'twice as much as' holds between two masses iff $m_a = m_b/2$. And, significantly, *any* representation of these fundamental mass relations satisfies TMR. If this were not the case — if TMR were true, say, on the kilogram scale, but false in pounds — then we would not consider it an objective fact. Therefore, what explains TMR on the literalist account is that mass properties stand in certain relations to each other and that however we represent this structure numerically, for

[6] If it seems as if these second-order relations still sneak in Platonic numbers, we can replace them with non-numerical relations of between-ness and concatenation; see Eddon (2013) for further details.

any three masses m_a, m_b and m_c it is true that $\frac{m_i}{m_k} = \frac{m_i}{m_j} \cdot \frac{m_j}{m_k}$.

The above explanation only appeals to the relational structure of mass magnitudes, and neither to intrinsic numerical properties nor to intrinsic identities. Hence, the conventionalist can give exactly the same explanation of TMR; after all, the conventionalist too believes that mass properties have a relational structure which we can represent numerically. She differs from the literalist in that she repudiates intrinsic identities, but intrinsic identities of mass magnitudes play no role in explaining TMR. Or, to put it differently: the literalist and the conventionalist disagree over which assignments of objects to mass values represent distinct physical possibilities; but whether or not such assignments are physically equivalent is entirely orthogonal to whether such assignments violate TMR. Therefore, the conventionalist can explain the transitivity of mass ratios equally well as the literalist.

However, have we not given in to comparativism by now, explaining the conspiracy of mass ratios by an appeal to fundamental relations between mass properties? The answer is no, because these relations between mass properties are quite different from the mass relations between *objects* which the comparativist posits. In fact *any* explanation of a quantitative relation in terms of the properties of its relata requires an appeal to the relations between these properties. Suppose, for example, that we want to explain why the Eiffel Tower is taller than the Taj Mahal. We start with noting that the Eifel Tower is 324 m, and the Taj Mahal 73 m. But this is not enough: we also need the fact that 324 m is *taller* than 73 m, i.e. a relation between height properties. Usually, these facts are so obvious to us that we need not mention them, but a proper explanation of the relations between objects in terms of their intrinsic properties requires an appeal to relations between these properties. Nevertheless, this does not make us relationists (or comparativists) in the sense that relations between *objects* are fundamental, rather than relations between properties. Therefore, the conventionalist can explain the conspiracy of mass relations without collapsing into relationism.

4.5 Conclusion

We have now explored three views on the nature of masses. Firstly, there is literalism about intrinsic mass properties: intrinsic mass properties with intrinsic identities are the fundamental mass magnitudes. According to this view, models related by mass scalings represent distinct physical possibilities, and the problem with this is that the differences between these worlds are in principle undetectable. Secondly, there is literalism about mass relations, or comparativism: mass *relations* (with intrinsic identities) are the fundamental mass magnitudes. On this view, models related by mass scalings are physically equivalent, since they agree on all mass relations. However, the literalist about mass relations has to posit cosmic coincidences in the instantiation pattern of mass relations by pairs of objects. Finally, I have set out a novel position: *conventionalism* about intrinsic mass properties. On this view, intrinsic mass properties *without* intrinsic identities are the fundamental mass magnitudes. I have argued that this view *avoids* vicious undetectability, since it identifies symmetry-related models, and *explains* the transitivity of mass ratios, hence avoiding charges of conspiracy.[7]

Therefore, conventionalism about mass properties is the best option. And from this it follows that there are physically real non-invariant quantities, *contra* the Invariance Principle. For according to the above argument intrinsic masses, which are not invariant under mass scalings, are explanatory: positing real mass properties explains the conspiracy of mass ratios. And, according to (1) of the Inconsistent Triad, a quantity which is explanatory is physically real. Therefore, intrinsic masses are physically real.

In the next two chapters, I will apply these same arguments to quantities which are not invariant under *gauge* symmetries using the Aharonov-Bohm effect as a case study. I will conclude that the vector potential field is physically real, and hence that (2) of the Inconsistent Triad is false.

[7]What about the natural fourth option, conventionalism about mass relations? If there are no further symmetries under which mass relations are non-invariant, this view is practically identical to literalism about mass relations, since the intrinsic identities then make no difference to which instantiation patterns are possible.

Chapter 5

The Aharonov-Bohm Effect

We have seen in Chapter 1 how the Inconsistent Triad prompts an investigation into the ontological status of gauge quantities based on their explanatory potential. Chapter 2 sketched a conventionalist approach to quantities according to which assignments of values are partly conventional. In Chapter 3 we saw that the main arguments against the reality of gauge quantities failed once this conventionalist account was adopted. Finally, in Chapter 4 a case was made for the explanatory power of non-invariant quantities, using the example of mass. It follows that the realist should interpret such quantities as physically real.

However, I have not yet delivered on the promise of this thesis, since so far I have only considered quantities which vary under *global* symmetries. In this and the next chapter, I will apply the conventionalist account to quantities which vary under *local* symmetries.

In this chapter, I will give an account of Aharonov-Bohm effect and survey the main responses. The main impact of the Aharonov-Bohm effect, which I explain in §5.1, is that it makes the standard interpretation of electrodynamics as involving electric and magnetic fields untenable: a new ontology is needed. In §5.2 I survey the main responses to the Aharonov-Bohm effect. The positions here are analogous to (literalist) absolutism and relationism, with the associated problems of indeterminism and cosmic coincidences respectively.

In the next chapter, I give a conventionalist account of gauge quantities. This is an extension of the Quantity-Value Conventionalism which yields the novel conclusion that for gauge fields there are no facts about cross-point value comparisons *within* the same possible world. I will then argue that this account can explain the cosmic coincidences which are pos-

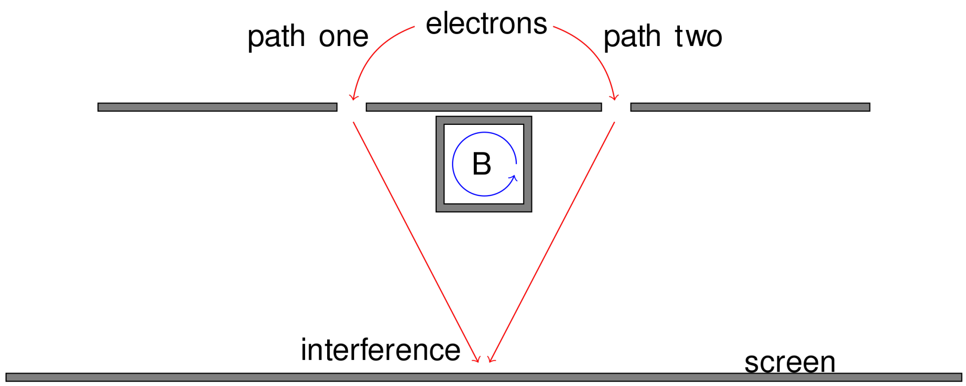

Figure 5.1: The set-up for the Aharonov-Bohm effect.

tulated as brute facts in the non-separable 'relationist' account propounded by Richard Healey. Therefore, we should admit the vector potential field into the ontology of semi-classical electrodynamics.

5.1 The Aharonov-Bohm Effect

In the below analysis of the Aharonov-Bohm effect, I closely follow a set of notes by Harvey Brown (2016). I will first sketch the general set-up and then work through a derivation of the effect.

5.1.1 Set-Up

The Aharonov-Bohm effect is essentially a modified double-slit experiment. As usual, we have an electron source with a plate with two slits placed in front of it, and a screen on the other side. But the Aharonov-Bohm effect adds the following twist: we place a solenoid just behind the plate, such that its interaction with the electron wavefunction is negligible (see Fig 5.1). Without the solenoid, the double-slit experiment produces an interference pattern on the screen. The surprising result of the Aharonov-Bohm effect is that this interference pattern *changes* when we let a current run through the solenoid. Specifically, the positions of the interference peaks are *shifted* within the interference envelope.

What is surprising about this result is that the value of the magnetic field is *zero* outside the solenoid. Hence, we cannot simply understand the change in the interference pattern as the result of the magnetic field acting on the electron wavefunction. This led Aharonov and Bohm to posit the *vector* *potential field*, which does *not* vanish outside the solenoid, as physically real and causally responsible for the effect (Aharonov and Bohm 1959). However, according to the Invariance Principle gauge dependent quantities are physically unreal, and this has led philosophers to search for different explanations of the Aharonov-Bohm effect. I will consider some of these accounts in the next section.

5.1.2 Derivation

A derivation of the Aharonov-Bohm effect can aspire to various levels of sophistication. Some opt for the full arsenal of the fibre bundle formalism, but I agree with Wallace (2014) that this level of detail is superfluous. I will also put things in terms of the three-vector potential field $\mathbf{A}$, rather than the four-vector A^{μ}.

I will start with a brief exposition of the gauge freedom of classical electromagnetism. I then give a derivation of phase shift in the Aharonov-Bohm effect. For subtler and more comprehensive treatments see Peshkin and Tonomura (1989); Healey (2007); Earman (2017).

Electromagnetic Potentials

In classical electrodynamics, the electric and magnetic fields can be expressed in terms of the *scalar potential* ϕ and the *vector potential* $\mathbf{A}$, which are defined as:

$$\mathbf{B} = \nabla \times \mathbf{A} \tag{5.1}$$

$$\mathbf{E} = -\partial_t \mathbf{A} - \nabla \phi \tag{5.2}$$

One reason for expressing $\mathbf{E}$ and $\mathbf{B}$ in terms of ϕ and $\mathbf{A}$ is that when we do, two of Maxwell's four equations are guaranteed to hold. Gauss' Law, for example, becomes a mathematical theorem: $\nabla \cdot \mathbf{B} = \nabla \cdot (\nabla \times \mathbf{A}) = 0$. The electromagnetic potentials are thus convenient mathematical tools.

Equations (5.1) and (5.2) are invariant under the following gauge transformation:

$$\mathbf{A} \to \mathbf{A} + \nabla f \tag{5.3}$$

$$\phi \to \phi - \partial_t f \tag{5.4}$$

where f is an arbitrary function of x. The fact that electrodynamics is invariant under the gauge transformations (5.3) and (5.4) has led most physicists to interpret $\mathbf{A}$ and ϕ as useful mathematical tools which don't represent anything physical, since if they did represent something physical solutions to Maxwell's equations related by gauge transformations would represent physically distinct but empirically indistinguishable possibilities. As we will see below, the Aharonov-Bohm effect puts this assumption into question.

Phase Shift

In order to derive the Aharonov-Bohm effect, start with the solenoid turned off. Here we may work in the 'natural gauge', $\mathbf{A}_0$, in which $\mathbf{A}$ is set to zero everywhere.

We can write down the electron wavefunction of the double-slit experiment as follows:

$$\psi^0(Q) = \psi_I^0(Q) + \psi_{II}^0(Q), \tag{5.5}$$

where ψ_I and ψ_{II} are the components of the wavefunction going through the left and right slit respectively, and Q is an arbitrary point on the projection screen. The superscript 0 denotes that we are working in the natural gauge.

How does ψ^0 transform under an arbitrary gauge transformation $\mathbf{A}_0 \rightarrow \nabla f$? We need to look at the Schrödinger equation to find out. The Hamiltonian of an electron moving through a magnetic field has the following form:

$$\mathcal{H} = \frac{1}{2m}(-i\nabla - q\mathbf{A})^2 \tag{5.6}$$

We can easily confirm that in order for the Schrödinger equation to be invariant under gauge transformations of the form (5.3), ψ must transform as follows:

$$\psi(x) \rightarrow \psi'(x) = e^{ief(x)}\psi(x) \tag{5.7}$$

Therefore, ψ and $\mathbf{A}$ together are invariant under the following set of gauge-transformations:

$$\mathbf{A} \rightarrow \mathbf{A} + \nabla f \tag{5.3}$$

$$\psi(x) \rightarrow e^{ief(x)}\psi(x) \tag{5.7}$$

Since f is an arbitrary function (subject to certain conditions such as smoothness), we can always pick an f such that $f(P) = 0$, where P is the location of the electron source. This enables us to rewrite $f(Q)$ as follows:

$$f(Q) = f(Q) - f(P) = \int_P^Q df = \int_P^Q \nabla f \cdot d\mathbf{x} = \int_P^Q \mathbf{A} \cdot d\mathbf{x} \tag{5.8}$$

Therefore, we can express the gauge transformation (5.7) of the wavefunction in terms of $\mathbf{A}$ as follows:

$$\psi(y) \to \psi'(y) = \exp\left(ie \int_P^y \mathbf{A} \cdot dx\right)\psi(y) \tag{5.9}$$

We now consider what happens when a solenoid is placed behind the screen. Whether or not the solenoid is turned on, the magnetic field is zero along any arbitrary path from the source to the screen, assuming that the electron does not penetrate the solenoid. Therefore, we can use the gauge transformation (5.9) for the first term of the double-slit wavefunction:

$$\psi_I^0(Q) \to \exp\left(ie \int_P^Q \mathbf{A} \cdot d\mathbf{x}_I\right)\psi_I^0(Q), \tag{5.10}$$

where $d\mathbf{x}_I$ is an arbitrary path from P to Q through the left slit. Of course, ψ_{II}^0 will undergo a similar transformation. Therefore, the total wavefunction transforms as follows:

$$\psi^0(Q) \to \exp\left(ie \int_P^Q \mathbf{A} \cdot dx_I\right)\psi_I^0(Q) + \exp\left(ie \int_P^Q \mathbf{A} \cdot dx_{II}\right)\psi_I I^0(Q) \tag{5.11}$$

$$= \exp\left(ie \int_P^Q \mathbf{A} \cdot dx_I\right)\left[\psi_I^0(Q) + \exp\left(ie \oint \mathbf{A} \cdot dx\right)\psi_{II}^0(Q)\right] \tag{5.12}$$

We can ignore the global phase and simply write this as:

$$\psi(Q) = \psi_I^0(Q) + \exp\left(ie \oint \mathbf{A} \cdot dx\right)\psi_{II}^0(Q) \tag{5.13}$$

Using Stokes' theorem, we can rewrite the loop integral to obtain:

$$\oint \mathbf{A} \cdot dx = \iint (\nabla \times \mathbf{A}) \cdot \mathbf{n} dS = \iint \mathbf{B} \cdot \mathbf{n} dS = \Phi \tag{5.14}$$

where Φ is the flux through the surface bounded by the loop. This gives us the following final expression for the electron wavefunction in the Aharonov-Bohm effect:

$$\psi(Q) = \psi_I^0(Q) + \exp(ie\Phi)\ \psi_{II}^0(Q) \tag{5.15}$$

Note that (5.15) is a gauge-*independent* expression, since Φ depends on **B** rather than on **A**! If Φ is zero — that is, if there is no current running through the solenoid — the phase difference vanishes, and we obtain the usual wavefunction for the double-slit experiment. However, when we turn on the solenoid there is a non-zero phase difference, which causes the interference pattern to shift as described above (unless $\Phi = 2\pi n$). This, then, is the puzzling result of the Aharonov-Bohm effect: a magnetic field seems to act on an electron which moves through an area in which that field is zero!

5.2 Interpreting Electromagnetism

The Aharonov-Bohm effect has led to a number of alternative interpretations of electro-dynamics. In this section, I will survey the main approaches: keeping the **B**-field; a realist interpretation of the **A**-field; positing non-localised *loop holonomies*; or Wallace's (2014) novel pair of gauge-invariant quantities.[1] We will find that each of them has serious drawbacks. In the next chapter, I will add my own *conventionalist* account of the Aharonov-Bohm effect to this list and argue that it offers the greatest explanatory strength without serious disadvantages.

5.2.1 Magnetic Field

The first option is to stick with electromagnetism as usually interpreted: the electromagnetic field is physically real, and the potential fields are only mathematical devices. The problem with this approach is that it implies action at a distance: because the magnetic field is zero outside the solenoid, where the amplitude of the electron wavefunction is non-negligible, it does not interact with the latter locally. Therefore, the magnetic field *inside* must act on the electron *at a distance*.

As far as I am aware, no one has seriously defended this interpretation of electrodynamics since the prediction of the Aharonov-Bohm effect, especially once the effect was empirically confirmed. Giving up the principle of local

[1] It is not my aim to give a comprehensive overview of the literature, and therefore I don't discuss other interpretations, including those of Leeds (1999); Nounou (2003); Mattingly (2006); Weatherall (2014).

action is a cost no one is willing to pay. I will therefore not consider this approach here.

5.2.2 Vector Potential

The seemingly intuitive response to the Aharonov-Bohm effect is to reify the vector potential field. After all, **A** is *not* zero outside the solenoid, and when we integrate **A** along the path of the electron wavefunction we get exactly the Aharonov-Bohm phase difference back. Furthermore, it is a **A** and not **B** which appears in the Hamiltonian for a charged particle moving through a magnetic field (Eq. 5.6), which suggests that it is the vector potential field which dynamically interacts with the electron.

This interpretative option has long been popular with physicists. Aharonov and Bohm themselves wrote that:

> [I]n a theory involving only local interactions [...] the potentials must, in certain cases, be considered as physically effective, even when there are no fields acting on the charged particles. (Aharonov and Bohm 1959, 490)

A few years later, Feynman simply *defined* what it means for a field to be real in terms of whether it acts locally:

> In our sense then, the A-field is "real." You may say: "But there *was* a magnetic field." There was, but remember our original idea—that a field is "real" if it is what must be specified *at the position* of the particle in order to get the motion. The B-field in the whisker acts at a distance. If we want to describe its influence not as action-at-a-distance, we must use the vector potential. (Feynman et al. 1964)

But Feynman too easily dismisses the gauge-variance of the vector potential which philosophers find so problematic. As we saw in Chapter 3, interpreting non-invariant quantities as physically real can lead to issues such as indeterminism and undetectability. For example, for any solution to Maxwell's equations we can construct another solution which agrees on all values of **A** before some time t, but disagrees on some of the values of **A** after t. Since these solutions disagree over the strength of the vector potential field at some points in spacetime, it seems that they represent physically distinct possibilities. Thus, on this interpretation electrodynamics is indeterministic: the whole history of the universe up to time t does not fix what happens after t. Furthermore, gauge-related models are empirically indistinguishable from each other. Therefore, the values of $\mathbf{A}$ are in principle undetectable.

Maudlin (1998) calls this option the ONE TRUE GAUGE interpretation. ONE TRUE GAUGE is a literalist-absolutist interpretation of the vector potential field, since it implies that gauge-related models are physically distinct. It is therefore not surprising that it leads to indeterminism and undetectability; these issues also arose for absolutist interpretations of location and mass. In Chapter 3, I argued that a conventionalist account of these quantities allows us to avoid these vices. The same is true in the case of gauge: if we can accept that gauge-related models are physically equivalent representations of the same field distribution, we can be realists about the vector potential field without falling prey to indeterminism. In the next chapter, I detail the conventionalist approach on which this is the case.

5.2.3 Holonomies

The $\mathbf{A}$-field is not gauge-invariant, but the so-called *loop holonomies* are:

$$\exp\big[-i\oint \mathbf{A}\cdot dx\big]. \tag{5.16}$$

Since the integral of $\mathbf{A}$ around a closed loop is equivalent to Φ through the enclosed area, holonomies are gauge-invariant quantities of electrodynamics. Furthermore, Wu and Yang (1975) have shown that the holonomies provide a *complete* specification of the gauge-invariant content of electrodynamics: if we know the holonomies attaching to all space-time loops, then we know all the gauge-invariant facts. This is also evident from the Aharonov-Bohm effect, since the phase difference measured there is proportional to the loop holonomy of the closed path of the electron wavefunction.

This has led some to argue that the holonomies represent the fundamental subject matter of electrodynamics (Belot 1998, 2003; Lyre 2004; Healey 1997, 2007). On this picture, there is a fundamental property associated to every closed curve in spacetime which is *non-localised*: it is not composed of field-values at individual space-time points, but attaches to the curve as a whole. When a charged particle moves along a closed path in spacetime, it interacts with the holonomy 'at once' around the whole loop. If the particle does not complete a loop — for example, if we intercept the electron of the

Aharonov-Bohm effect before it passes the solenoid — then no interaction takes place at all.

An ontology of loop holonomies is *non-separable*: the intrinsic facts about a spacetime region $X \cup Y$ do not supervene on the intrinsic facts about X on its own and the intrinsic facts about Y on its own. We can easily see this when we consider two partially overlapping regions X and Y close to the solenoid, as in the following diagram:

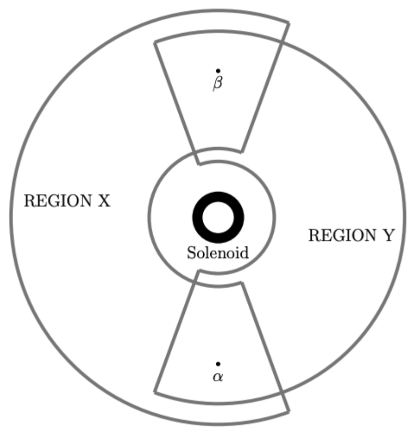

Figure 5.2: The holonomies in X and Y are zero, but there are non-trivial holonomies around the region $X \cup Y$ which encloses the solenoid. Reproduced from Wallace (2014).

Recall that the holonomy of a closed path is equal to the flux through the surface which the path encloses. Since X does not enclose the solenoid, the flux through the surface it encloses is zero, and similarly for Y. But now consider the union of X and Y, $X \cup Y$. This region *does* enclose the solenoid and thus has a non-zero holonomy value. We could not have derived this non-zero value from the zero-valued holonomies of regions X and Y. Therefore, the intrinsic facts about X and Y on their own do not determine the intrinsic facts about $X \cup Y$, and hence separability fails.

We already encountered a non-separable ontology in the previous chapter: comparativism about mass. There, mass relations between distant objects did not supervene on their intrinsic properties. There is nothing objectionable about non-separability *per se*. Indeed, quantum mechanics gives us good reasons to believe that the world *is* fundamentally non-separable. But the non-separability of loop holonomies implies that Healey's account

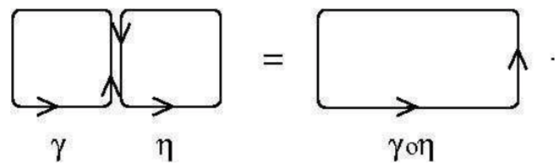

Figure 5.3: The loop $\gamma \circ \eta$ is the concatenation of γ and η. Reproduced from Asselmeyer-Maluga (2016).

is a version of *relationism*: its fundamental ontology consists of polyadic relations between sets of points which make up spacetime loops. As we saw in Chapter 4, relationist ontologies can lead to cosmic coincidences, and this is also the case here. Firstly, there is an analogue to the conspiracy of mass ratios, which we may call the *conspiracy of holonomy relations*: the proponent of holonomies has to postulate brute relations between holonomies. Secondly, the holonomy account of the Aharonov-Bohm effect violates the principle of local action. According to Healey, the holonomy interacts with the wavefunction 'at once' over its entire path; therefore, this account does not only involve non-separable fundamental properties, but also non-separable *causal action*. This non-separable causation is a form of action at a distance.

The Conspiracy of Holonomy Relations

Recall that the comparativist about mass had to posit a constraint on the fundamental mass relations: the transitivity of mass relations (TMR). Without this constraint, we cannot represent mass relations as mass ratios. In the previous chapter, I argued that TMR is either a cosmic coincidence, or enforced through action at a distance.

A similar dilemma occurs for the holonomy interpretation of electrodynamics. The holonomies of two loops l_1 and l_2 satisfy the following relation:

$$H(l_1 \circ l_2) = H(l_1)\, H(l_2) \tag{5.17}$$

Here, $l_1 \circ l_2$ is the concatenation of the two loops, that is, the result of first going around l_1 and then going around l_2 (see Fig. 5.3). Call this feature *composite loop multiplication* (CLM).

CLM guarantees that we can represent holonomies as gauge fields, and hence express $\mathbf{B}$ in terms of $\mathbf{A}$. As Arntzenius puts it, "a fairly obvious explanation of why [CLM] hold[s] is that the map H is, roughly speaking, the integration of a connection around a loop". In other words, if we define $H(l) = \exp\left[-i/\hbar \oint_l \mathbf{A} \cdot d\mathbf{x}\right]$, then

$$\begin{aligned} H(l_1)\, H(l_2) &= \exp\left[-i \oint_{l_1} \mathbf{A} \cdot d\mathbf{x}\right] \cdot \exp\left[-i \oint_{l_2} \mathbf{A} \cdot d\mathbf{x}\right] \\ &= \exp\left[-i \oint_{l_1 \circ l_2} \mathbf{A} \cdot d\mathbf{x}\right] \\ &= H(l_1 \circ l_2) \end{aligned} \tag{5.18}$$

The second line follows from the first because we have to 'backtrack' when first integrating over l_1 and then over l_2, leaving us with the integral over $l_1 \circ l_2$. Therefore, we can explain CLM if we posit the existence of a local vector potential field.

On the holonomy interpretation, on the other hand, there is nothing which guarantees that CLM holds. For example, $H(l_1)$ could have been slightly lower than it actually is. In that case, $H(l_1 \circ l_2)$ would either have been different or would remain the same. If the former is the case CLM could easily have failed, and it seems a conspiracy that the holonomies just happen to be related in such a way that we can represent them as loop integrals of local field values. Again, there are again good reasons to believe that CLM *could* easily have failed. It is certainly true that the value of $H(l_1)$ easily could have differed, since there are dynamically possible models on which this is so. Furthermore, the world with a different $H(l_1)$-value closest to the actual world is one in which all other holonomies are the same, since this maximises the overlap of particular facts while minimising the violation of CLM to a single holonomy. The No Miracles Argument then implies that there *are* localised fields of which holonomies are composed.

Alternatively, $H(l_1 \circ l_2)$ would change if $H(l_1)$ were to change. But according to Lewis' definition of causal action, this implies that $H(l_1)$ causally acts on $H(l_1 \circ l_2)$. Since the latter loop does not overlap the former, this is similar to action at a distance. Of course, it is true that l_1 and $l_1 \circ l_2$ *partially* overlap. But it is not possible to argue that $H(l_1)$ simply acts on that part of $H(l_1 \circ l_2)$ with which it overlaps, since holonomies are fundamental *non-localised* properties: they are *not* composed of properties of open paths. Therefore, there is no property which is located at the overlap of l_1 and $l_1 \circ l_2$, and $H(l_1)$ has to act on $H(l_1 \circ l_2)$ as a whole.

Healey (2007, 4.4.2) gives a few responses to this objection. Firstly, he claims that using $\mathbf{A}$ in an explanation is illegitimate since $\mathbf{A}$ is undetectable.

But Healey offers no reason to believe that undetectable structure cannot explain. Of course, he may think that the cost of introducing undetectable quantities is too high *even if* such quantities are explanatory, but in the next chapter I show that it is possible to give a realist interpretation on which $\mathbf{A}$ is not undetectable (see also §3.3). Secondly, Healey argued that unlike most explanatory quantities, $\mathbf{A}$ has no unifying power. However, unification is far from the *only* explanatory virtue; making sense of cosmic coincidences is another, and so it is not the case that the lack of unifying power counts against $\mathbf{A}$.

The third response is more challenging. Healey appeals to the *loop supervenience* of holonomy properties: "the holonomy properties of any loop $\otimes_i L_i$ are determined by those of any loops L_i that compose it" (Healey 2007, 123). This means that the composite loop is *not* fundamental, since it is composed of smaller loops. Of course, the same is true for these smaller loops, which are composed of even smaller loops. There is no natural stopping point, and hence no smallest fundamental unit. But let's accept this for now. The question then is: can loop supervenience avoid the dilemma between cosmic coincidences and action at a distance?

Firstly, note that the sense in which small loops 'compose' larger loops is not straightforward. It is not the case that the smaller loops make up larger loops in the sense in which the masses of two hemispheres make up the sphere's total mass. In the latter case, we are simply talking about mereological composition; but large loops are not the mereological composite of small loops, since the small loops contain parts which do not overlap with any part of the large loop (see Fig. 6.3). Furthermore, loop composition does not have the right formal properties to count as mereological fusion. For example, mereological part-hood is asymmetric, since objects cannot become parts of each other. But we can prove that loop composition *is* symmetric. Let l_1^{-1} stand for the loop which goes around l_1 in the opposite direction. Then $H(l_1 \circ l_2)\, H(l_1^{-1}) = H(l_2)$. If loop supervenience is identical to mereological fusion, then this implies that $l_1 \circ l_2$ is part of l_2. But since l_2 is also part of $l_1 \circ l_2$, this means that loop supervenience is a symmetric relation and so cannot be identical to mereological composition. But if the supervenience of loops on each other has no such metaphysical basis (and no suggestions are offered), then Healey's response is nothing more than a question-begging restatement of the very cosmic coincidence we are trying to explain.

Secondly, note that loop holonomies are causally efficacious since they act on charged particles. Suppose an electron goes around the loop $l_1 \circ l_2$, and hence is acted on by $H(l_1 \circ l_2)$. If $H(l_1 \circ l_2)$ supervenes on $H(l_1)$

and $H(l_2)$, then likewise the action of $H(l_1 \circ l_2)$ supervenes on the action of $H(l_1)$ and $H(l_2)$ on the particle. However, the latter two loops do not completely overlap the electron's trajectory, and so their influence on the electron violates the principle of local action. Again, it is impossible to argue that the parts of $H(l_1)$ and $H(l_2)$ which *do* overlap the electron's trajectory act on it, since holonomies are non-localised properties which do not have parts. Therefore the appeal to loop supervenience, if it is coherent, cannot avoid action at a distance.

Non-Separable Causation

There is a further problem associated with the action of loop holonomies on charged particles. Let's call loops which are located at a single instant of time *space-like*, and loops which span across multiple instants *time-like* (for simplicity, assume we have an objective standard of simultaneity). According to Healey, space-like loop holonomies act on non-localised properties of the electron wavefunction, and this interaction is local since the holonomies act on properties of the wavefunction with which they are co-located. As Healey admits, "[th]ere is still an issue about just what these non-localized phase properties are" (128), and it is notoriously difficult to express the Hamiltonian in terms of holonomies, rather than field values. But even if we ignore these issues, there are serious doubts as to whether the proposal works.

The problem is that in some circumstances we must consider the holonomies of *time-like* loops as acting on the wavefunction. Nounou (2010) gives two reasons for thinking that time-like loops are involved in causal interactions. Firstly, she shows that the derivation of the Aharonov-Bohm effect I presented above is an approximation. The full solution involves loops which wind around the solenoid multiple times, in addition to simple closed loops. These 'winding' loops are time-like. Nounou (2010, 397-402) goes on to argue that it is impossible to reduce these winding loops to space-like loops without violating local action in the process. Secondly, we have only considered constant magnetic fields. But if we vary $\mathbf{B}$ with time, the phase difference of the Aharonov-Bohm effect is equivalent to the holonomy of the time-like loop which maps the electron's trajectory. This makes sense if we think of the electron as moving through a localised field, such that it is acted on by the field at every spacetime point of its trajectory. But there are no space-time loops which result in the same interaction. Suppose, for example, that we turn on the solenoid instantaneously when the electron is halfway between the slits and the screen. The space-like loops around the solenoid then either have holonomies of value 1 (when the solenoid is turned of) or of value $e^{-i\Phi}$ (when the solenoid is on). The time-like loop along which the trajectory of the electron lies, however, has a value in between these two since the surface it encloses cuts across the flux of the solenoid. Therefore, it is the holonomy of the time-like loop which acts on the electron.

What is the problem with action along time-like loops? If interactions happen over time-like loops then we cannot define instantaneous stages of this interaction. For example, there is no fact of the matter about the electron's non-localised phase before it has hit the screen, since its non-separable interaction with the holonomy is non-locally distributed over time. What this means is that interaction with time-like loops is not a *continuous* process, since it has no well-defined continuous succession of states; it happens 'at once' over the whole loop.[2] This violates Healey's principle of local action, which states that "if an external influence on A is to have any effect on B, that effect must *propagate* from A to B via some *continuous* physical process". Since the process through which time-like holonomies influence electrons is not continuous, it is not a form of local action.

An ontology of localised potentials can avoid both of these issues. First, as has been shown above, we can explain the conspiracy of holonomy relations by positing that holonomies are composed of local field values. And since these fields are localised in spacetime, their action on the wavefunction is local too: it is simply the pointwise interaction which is described by the Hamiltonian. Therefore, a field interpretation is preferable to a holonomy interpretation, all else equal. Of course, not all else is equal, since a literalist field interpretation implies indeterminism and undetectability. In the next chapter, I show that a conventionalist field interpretation has the same explanatory strength but avoids these vices.

5.2.4 Wallace's Proposal

Finally, I will discuss an ingenious recent proposal from David Wallace (2014). Wallace introduces two novel fields: ρ, which is the radial part of the wavefunction ψ, and $D\theta = \nabla\theta - \mathbf{A}$, where θ is the angular part of the wavefunction. Both quantities are gauge-invariant: ρ is a gauge-invariant feature of ψ, and $D\theta$ is gauge-invariant since transformations of $\mathbf{A}$ and θ cancel each other. In addition, both fields are localised. Finally, Wallace shows that we can rewrite the Schrödinger equation in terms of ρ and $D\theta$. We thus have a representation which is both local and gauge-invariant, and hence separable, deterministic and detectable. What more can we ask for?

[2]See Healey's (1994) definition of continuous processes (342).

There is a problem with Wallace's proposal: it does not generalise to non-Abelian gauge theories. Wallace (2014, 17) himself admits this is a problem, writing that "in general, I know of no comparably simple set of local gauge-invariant quantities in the non-Abelian case that can serve as a gauge-invariant representation".

Furthermore, even more complex *Abelian* theories have no *unique* gauge-invariant representations. Consider, for example, two complex-valued fields ψ and χ of different masses, both of which are coupled to $\mathbf{A}$. The combined Lagrangian is:

$$\mathcal{L} = \mathcal{L}_1 + \mathcal{L}_2 \tag{5.19}$$

where

$$\mathcal{L}_1 = (\partial^\mu + ie_1 A^\mu)\psi^\star(\partial_\mu - ie_1 A_\mu)\psi \tag{5.20}$$
$$\mathcal{L}_2 = (\partial^\mu + ie_2 A^\mu)\chi^\star(\partial_\mu - ie_2 A_\mu)\chi \tag{5.21}$$

$\mathcal{L}_1$ is just the Lagrangian of a single complex field coupled to the gauge field. Therefore, we can rewrite it in terms of the gauge-invariant quantities ρ_1 (the magnitude of ψ) and $D\theta_1^\mu = A^\mu - \partial^\mu\theta_1$, where θ_1 is the phase of ψ. But since we have 'replaced' A^μ with $D\theta_1^\mu$ in $\mathcal{L}_1$, we have to make the same substitution in $\mathcal{L}_2$. This gives us an ontology consisting of a real-valued field ρ_1 with charge e_1, a complex-valued field χ with charge e_2, and a connection $D\theta_1^\mu$. Since we now have a real-valued instead of a complex-valued field, we have reduced the number of degrees of freedom by one. As it turns out, this is enough to yield a gauge-invariant representation.

However, we could also have started with $\mathcal{L}_2$ and written *that* Lagrangian in terms of a real-valued field and a connection. This results in an ontology consisting of a complex-valued field ψ with charge e_1, a real-valued field ρ_2 with charge e_2, and a connection $D\theta_2^\mu$. But these are different ontologies! Although both consist of a real-valued field and a complex-valued field, the charges of these fields differ. On the first interpretation the real field has a charge e_1, but on the second its charge is e_2, and vice versa for the complex field.

Therefore, using Wallace's approach on spinor fields leads to a form of theoretical underdetermination. The choice between these two ontologies is arbitrary, since they are empirically equivalent. This is not much better than the situation in which there are multiple indistinguishable gauges from which one can choose.

All this suggest that it is a not more than a fortunate accident that we can represent the simple $U(1)$ gauge theory we have considered so far in terms of a unique set of gauge-invariant quantities. This simplified theory is unrealistic in practice, and so Wallace's proposal does not offer us an approach to interpreting gauge theories in general. On the other hand, the conventionalist account readily generalises, since we can always stipulate that gauge properties are qualitatively individuated and hence that gauge-related models are physically equivalent. I will discuss this in more detail in the next chapter, but note that Wallace essentially suggests the same approach for those cases for which his proposal fails: "in general [...] such objects tend to lack simple geometric descriptions and must be described, at the cost of some arbitrariness, by a particular choice of coordinatisation: in the case of gauge fields, by a particular choice of gauge". Overall, then, we should prefer the approach which consistently works, rather than Wallace's account which happens to work in the simple $U(1)$ case.

5.3 Conclusion

In this chapter I introduced the Aharonov-Bohm effect and surveyed the main responses to it. The main contenders are realism about the vector potential and realism about loop holonomies. These positions are analogous to absolutism and relationism in the case of space and mass, with similar drawbacks: the 'ONE TRUE GAUGE' interpretation leads to indeterminism and the postulation of undetectable values, whereas Healey's relationist interpretation implies a conspiracy of holonomy relations.

We need an intermediate position on which the vector potential field *is* real and thus explains the conspiracy of holonomy relations, but on which gauge-related models are *not* physically distinct and hence no indeterminism arises. In the next chapter, I will argue that conventionalism about gauge quantities is exactly the middle ground we are looking for.

Chapter 6

Towards a Conventionalist Account of Gauge Quantities

In this chapter I will give a conventionalist account of the Aharonov-Bohm effect. This account is an alternative to absolutism and relationism: it interprets the potential field as physically real but identifies gauge-related models. This enables us to both explain the conspiracy of holonomy relations and avoid the vicious indeterminism which is the consequence of a literalist interpretation of the vector potential field.

There are some subtle implications of the conventionalist account when we consider gauge quantities. Firstly, a consequence of the fact gauge symmetries are *local* symmetries is that there is no well-defined notion of identity across spacetime points; every spacetime point has its own *sui generis* set of gauge properties (§6.1). This implies what I call the *Spatial Exclusion Problem*, but as I argue in §6.2 this problem is not as worrisome as it seems.

Secondly, I will look at the interaction which takes place in the Aharonov-Bohm effect. On a literalist picture, we can track precisely where the wavefunction picks up a phase. On the conventionalist picture, this is not the case: there is no fact of the matter about *where* the phase difference comes about (§6.3). I will argue that we can nevertheless give a local account of the Aharonov-Bohm effect (§6.4).

6.1 Conventionalism About Gauge

The conventionalist account of gauge quantities is an extension of the account I developed in Chapter 2. The main ideas generalise as follows:

1. Gauge quantities are physically real even if their values do not consistently refer (QVC);
2. Gauge symmetries relate physically equivalent models (Leibniz Equivalence);
3. Gauge properties have no intrinsic identities (Anti-Quidditism).

Nevertheless, gauge quantities *are* different. The difference is that gauge quantities transform under *local* symmetries, rather than the global symmetries we have been considering so far. In the case of mass, for example, we considered transformations which uniformly scale masses by the same factor everywhere in space. In contrast, gauge symmetries are spacetime dependent: a gauge transformation is analogous to scaling masses at different locations by different factors. Of course, local mass scalings are no symmetries of Newtonian mechanics, but the symmetries of gauge theories *are* local in this sense. For example, the gauge freedom of electromagnetism allows us to add an arbitrary gradient $\nabla f(x)$ to the vector potential $\mathbf{A}$, which is a local transformation since f is a function of the spacetime coordinates.

The consequence of this is that *there is no longer a notion of identity between the gauge properties of distant points.*[1] We can prove this as follows. Suppose for reduction that there *is* a meaningful relation of identity between the field strengths of $\mathbf{A}$ at distant points. Let $\mathbf{A}_x$ stand for the value of the vector potential field at x. We can then make claims of the form '$\mathbf{A}_x = \mathbf{A}_y$' or '$\mathbf{A}_x \neq \mathbf{A}_y$'. But these claims are not invariant under gauge symmetries: since gauge symmetries are local, there are gauge transformations which change the value of $\mathbf{A}_x$ while leaving $\mathbf{A}_y$ the same. Therefore, if cross-point comparisons between field strengths are physically meaningful then gauge-related models are *not* physically equivalent, and this contradicts our assumption of Leibniz Equivalence. In other words, if gauge-related models are physically equivalent then there are no facts about whether gauge properties at distinct locations are identical.

Although the lack of cross-point identities sounds counterintuitive, there is no reason to believe that it poses a threat to understanding gauge quantities as representing something physically real. Firstly, as Maudlin (2007) points out, it is not the case that we can make *no* comparisons between the quantities at distant points, since the field values at distinct points x and y are still values of the *same* field. Specifically, both values are part of value spaces with the same geometric structure; for $\mathbf{A}$ and ψ this is the structure

[1] Maudlin (2007) also draws this conclusion.

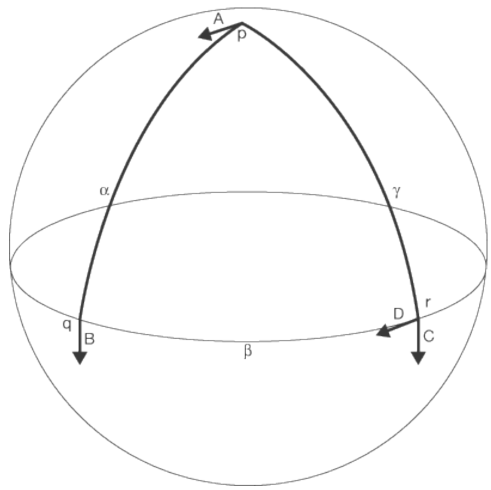

Figure 6.1: If we transport A along α and β we end up with C, but if we transport the same vector along γ we get D. Therefore, there is no fact of the matter about whether A and C point in the same direction. Reproduced from Maudlin (2007).

defined by the unitary group $U(1)$. As Maudlin (2007, 103) puts it, "there is no absolute comparison between physical states like color charge [or field strength] at different points, *but there are absolute identities of form at a more abstract level*" (emphasis mine).

Maudlin also discusses the familiar example of spatial direction in a curved space. If space is curved then there is no well-defined notion of sameness of direction for distant vectors, since each vector has its own tangent space. It is ill-defined, for example, whether the vectors labeled 'A' and ''C' in Fig. 6.1. point in the same direction on the 2D-surface of the sphere.[2] What we *can* do is parallel transport a vector along a particular path, i.e. move it around while keeping its orientation 'fixed'. If we start out with two vectors at different locations and parallel transport one towards the other, we can compare their directions when they meet. But this notion of sameness of direction is path-dependent, as Fig. 6.1 illustrates. If we parallel transport A from p to r via path γ the result is D, but if we take A from p to r via q, we end up with C instead. C and D point in different directions, so whether A points in the same direction as C or D depends on

[2]Of course, if we embed the sphere into a three-dimensional space we *can* compare these directions; but we are imagining that the sphere in Fig. 6.1 is not embedded in a higher-dimensional space.

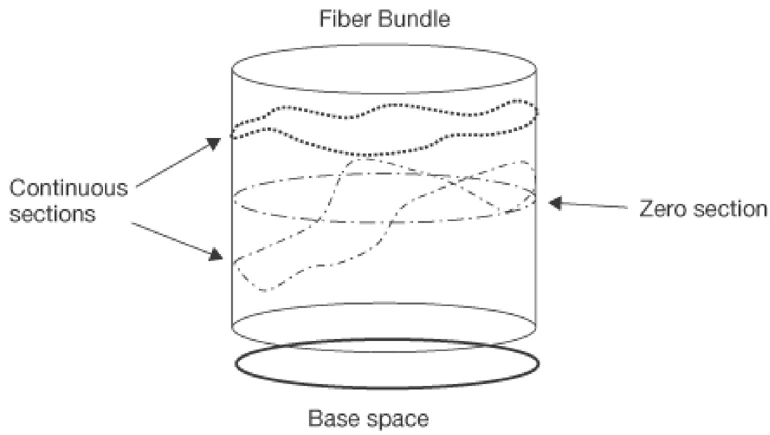

Figure 6.2: Reproduced from Maudlin (2007).

its path.

But of course this does *not* mean that spatial orientation does not exist. It is clear that each vector points in *some* direction. Rather, it means that direction is not a *universal*, since it makes no sense to claim that vectors at distinct locations have the same direction. The same conclusion applies to other gauge quantities, such as **A**: the fact that there are no cross-point comparisons between distant values of **A** does not mean that it cannot represent a physically real field. It only means that its field strengths are different from our intuitive conceptions of what fields are like.

Nevertheless, this conclusion is more radical than the anti-quidditism introduced in Chapter 2. According to anti-quidditism there are no meaningful cross-*world* comparisons between physical magnitudes. For example, there is no fact of the matter as to whether the mass of the Great Sphinx in the actual world is identical to the mass of the Great Sphinx' counterpart in a world with different mass ratios. We can conceptualise this as each world having its own *sui generis* set of mass properties. The conventionalist also advocates anti-quidditism about gauge quantities. For example, there are no cross-world comparisons possible between the values of **A**. But whereas it is possible to compare the masses of different objects *within* the same world, we have seen above that this is not the case for gauge quantities. The conventionalist account of gauge quantities thus goes beyond anti-quidditism.

We can understand this as each point in spacetime having its *own* set of field values, analogous to the idea that each possible world has its own set of mass values. Let $[\mathbf{A}]$ stand for space set of all values of $\mathbf{A}$. Instead of there being a single space $[\mathbf{A}]$ of field values which are instantiated at different points, on this picture each point x has a distinct space of field values $[\mathbf{A}]_x$. The elements of $[\mathbf{A}]_x$ are distinct from those of $[\mathbf{A}]_y$, and therefore x and y can never instantiate the exact same field value. In the language of fibre bundles, the value space which comprises $[\mathbf{A}]$ is called a *fibre*. There are copies of $[\mathbf{A}]$ at each point of the *base space*, which is spacetime. We refer to the local value space $[\mathbf{A}]_x$ as the *fibre at* x.[3]

Fig 6.2 illustrates this: it depicts a circular base space with a bundle of fibres. At each point of the base space there is a vertical fibre which contains values with a certain structure; for example, the fibres in Fig. 6.2 are isomorphic to the real numbers between 0 and 1. The fibres of $[\mathbf{A}]_x$ have the structure of an *affine space*, i.e. a vector space without an origin. The fibres also have a horizontal structure, which is equivalent to the structure of the base space. For example, there is a determinate 'distance' between the values of two distinct fibres. Note that these are not *actual* spatial distances, since field values are abstract properties living in an abstract value space. The function of the horizontal structure is to enable us to match up the value space of the vector potential field with physical space so that there is a unique fibre of field values at each spacetime point.

As I argued above, there are no comparisons between field values at distinct points. This means that there are no second-order relations between the pairs of values from distinct fibres. But at the moment our fibre bundle has *too* little structure, since we can assign numbers to its elements however we want. In order for the bundle to represent the vector potential field, we need to define second-order relations which are invariant under gauge transformations. Compare this with mass space: in order to represent masses numerically we had to define second-order mass relations between mass values. These relations represent invariant mass ratios. The invariant relations of $\mathbf{A}$ are properties of *loops* in value space. Define a *continuous section* as a set of points in the fibre bundle which form a continuous loop around the bundle's horizontal structure; see Fig. 6.2. We can assign each section a *second-order holonomy property*. Note that unlike the loop holonomies considered in the previous chapter, these holonomies are not properties of spacetime loops but of sets of field values. This is analogous to the second-order mass ratios: these are second-order relations between mass values rather than

[3] In canonical presentations of the fibre bundle formalism the vector potential field is a *connection* on the fibre bundle of ψ, rather than having its own fibre bundle. But there is nothing which prevents us from using the fibre bundle formalism to represent the value space of $\mathbf{A}$ instead, which is what I have done here.

fundamental relations between massive objects, as on the comparativist account. It is then possible to prove a *representation theorem* which states that the second-order holonomies determine a numerical representation of the fibre bundle which is unique up to gauge transformations (Healey 2007, §3.1.2). The values of $\mathbf{A}$ are then represented as three-dimensional vectors, and the second-order holonomies are identical to closed loop integrals over these values.

We can use this to show that gauge-related models are physically equivalent on this account. Consider a loop l in spacetime (rather than a section of the fibre bundle). Each point x on this loop instantiates a field value from the fibre $[\mathbf{A}]_x$. If we gather all the values instantiated around l, we obtain a continuous section of the fibre bundle (the 'lift' of l) which has a (second-order) holonomy property. If we gauge transform the values of $\mathbf{A}$ around l, we obtain another representation of the vector potential field. The question now is whether this is a representation of the same possible world. The answer is 'yes', since holonomies are invariant under gauge transformations. Therefore the section of the field values around l has the same holonomy property. But since there are no qualitative facts about field values other than their holonomy properties, these two gauge-related representations are qualitatively identical. Compare this again with the case of mass: if we double all mass values we obtain a novel representation. But the second-order mass ratio relations between the instantiated mass properties are the same on this representation, and hence the two are qualitatively identical. If field values have no quiddities, this means that these representations are physically equivalent.

It thus follows from the conventionalist account of gauge quantities I have sketched that gauge-related models are physically equivalent. This means that we can avoid indeterminism: the models which supposedly represent distinct possible futures for a given past are merely distinct representations of the same possible future. Similarly, this means that there are no undetectable differences between gauge-related models, since again these models are representations of the same state of affairs.

Furthermore, the conventionalist account of $\mathbf{A}$ allows us to explain Composite Loop Multiplication (CLM). Recall that CLM states that $H(l_1 \circ l_2) = H(l_1)\,H(l_2)$. If holonomies are intrinsic non-separable properties of spacetime loops, then CLM is a cosmic coincidence. However, if we define $H(l) = \oint \mathbf{A}\, d\mathbf{x}$ then CLM is a mathematical theorem (§5.2.3). On the conventionalist account there *are* local field values which we represent as three-dimensional vectors. Therefore it is legitimate to define $H(l)$ as the integral of $\mathbf{A}$ over l, and hence CLM is nomologically necessary. This means that the conventionalist account has a greater explanatory strength than the holonomy interpretation: it can explain relations which otherwise appear as conspiratorial brute facts. If all else is equal, realism about the vector potential field is thus preferable over fundamental loop holonomies. And all else *is* equal, since the conventionalist account avoids the theoretical vices associated with ONE TRUE GAUGE. I will now defend these claims against an objection.

6.2 The Spatial Exclusion Problem

On the conventionalist account of gauge quantities, each spacetime point x has its own set of field values $[\mathbf{A}]_x$: the fibre at x. But this fibre is an abstract value space, and I have given no explanation of what connects $[\mathbf{A}]_x$ and x to each other. It seems at least metaphysically possible that the field at x could take on a value from the fibre at y, i.e. $\mathbf{A}_x \in [\mathbf{A}]_y$. The fact that this does not happen — that, for any x, $\mathbf{A}_x \in [\mathbf{A}]_x$ — seems like a cosmic coincidence.

Call this the *Spatial Exclusion Problem.* It is a serious problem, since if conventionalism also has to posit a cosmic coincidence it has no advantage over the holonomy interpretation. I have no solution to the Spatial Exclusion Problem. However, I will argue in this section that there is an important disanalogy between the Spatial Exclusion Problem and the conspiracy of holonomy relations. As we saw in Chapter 5, the conspiracy of holonomy relations implies a violation of local action. This is not the case for the Spatial Exclusion Problem, and hence conventionalism is still preferable over relationism.

The conspiracy of holonomy relations concerned the relations between the properties of distinct loops in spacetime: loops appeared to coordinate their holonomy properties such that Composite Loop Multiplication was satisfied. But since a composite loop does not overlap with its 'parts', the fact that it nevertheless seems to 'know' its parts' values indicates a problematic form of instantaneous communication at a distance. The Spatial Exclusion Problem, on the other hand, concerns the instantiation patterns at individual spacetime points. There is no coordination *between* spacetime points to instantiate field values in a certain pattern. In other words, the field at x does not need to 'know' the values at y in order to restrict itself to $[\mathbf{A}]_x$, and there are no conspiratorial correlations between $\mathbf{A}_x$ and $\mathbf{A}_y$. Therefore, spatial exclusion does not violate the principle of local action.

The Spatial Exclusion Problem is similar to another problem in the metaphysics of quantities: what makes it the case that individuals instantiate at most *one* value of the same determinable? For example, why can one object not have two distinct masses? This problem is difficult to solve: Bricker (2017) considers five different solutions, none of which seem to work. *This* exclusion problem also applies to holonomies: there is no *a priori* reason to believe that a loop in spacetime cannot have multiple holonomy properties.

We thus have to posit inexplicable exclusions in any case, whether we are conventionalists or relationists. But if this is so, adding an additional exclusion principle is not as problematic. Perhaps it is a brute fact about nature that magnitudes of the same determinable exclude each other. Similarly, I will simply stipulate that field values at different fibres in the fibre bundle exclude each other without further explaining this fact.

6.3 The Aharonov-Bohm Effect (Again)

The holonomy interpretation of electrodynamics faces two problems: the conspiracy of holonomy relations and non-separable causation. I argued in §6.1 that conventionalism explains the conspiracy of holonomy relations, and it is now time to look at the interaction which causes the Aharonov-Bohm effect.

If we subscribe to a literalist interpretation of $\mathbf{A}$, explaining the Aharonov-Bohm effect is simple: the electron wavefunction just picks up a phase as it moves through the vector potential field. We can imagine the vector potential as a force which rotates the wavefunction of the particle in phase space. When we turn on the solenoid, we can track the phase of both parts of the wavefunction and see (for example) the left half picking up a greater phase than the right half as it moves through a region in which the $\mathbf{A}$-field is stronger. Furthermore, we can compare the phase of both parts of the wavefunction to each other at each instant. When the two parts meet each other at the screen, this phase difference gives rise to the Aharonov-Bohm effect.

The conventionalist account differs from the literalist's explanation on two points. Firstly, there are no cross-point comparisons between phase properties possible according to the conventionalist. This follows from the fact that ψ also varies under gauge transformations; the argument is analogous to the one I presented in §6.1. There is thus no fact of the matter about what the phase difference between the two halves of the wavefunction is before the electron hits the screen, and this means that we cannot track the phase difference build up. Secondly, we also cannot track how the phase of each individual part of the wavefunction changes over time, again because of the gauge freedom of semi-classical electrodynamics; here it is the integral of **A** over an open path which depends on our choice of gauge. As we will see, this means that there are no facts about *where* the wavefunction changes its phase.

Of course, the total phase difference around a closed loop *is* invariant, since the Aharonov-Bohm effect is measurable. But this does not help us to understand the Aharonov-Bohm effect as a *local* interaction between the vector potential field and the electron wavefunction, as I will argue we can. I will discuss both disanalogies in turn, arguing that neither prevents us from offering a local account of the Aharonov-Bohm effect.

In fact, we have already discussed the first point. The fact that there are no cross-point comparisons of phase properties in the Aharonov-Bohm effect is analogous to the impossibility of cross-point comparisons of spatial direction discussed above. Consider again the parallel transport of a vector around the earth. If we start at the north pole with two distinct vectors, we can transport them along different paths such that both point in different directions when meeting again at the equator. As we have seen, there are no meaningful and path-independent comparisons between the vector's directions along their journey. So we cannot track the divergence in their direction over time. As a result, there is no answer to the question *where* the 'phase' difference between the vectors arises. Nevertheless, we can clearly understand this topological analogue to the Aharonov-Bohm effect in a purely local fashion. Both vectors change their internal direction while traveling through space, and neither vector acts on the other at a distance.[4]

Similarly, we can understand the Aharonov-Bohm effect as a process in which there is no matter of fact as to *where* the phase difference comes about, but nevertheless there *is* a phase difference at the end of the process. This is perhaps puzzling, but it is no more puzzling than the examples discussed above. However, there is also an important difference. We can track the rotation of a vector as it moves through space, and similarly in the Twin Paradox each twin can measure her own proper time on a clock. In the Aharonov-Bohm effect, on the other hand, we cannot track how the phase of either part of the wavefunction changes as it moves through the vector potential field. This second point is subtle, and I will discuss it in the next section.

[4]The Twin Paradox has a similar structure. Due to the relativity of simultaneity, there are no meaningful facts about the age difference of the twins when they are apart. Nevertheless, we can understand the twins as picking up proper time while traveling along their respective worldlines, and no action at a distance is involved.

6.4 Interaction on the Conventionalist Account

I claim that the phase difference of the Aharonov-Bohm effect is the result of **A**'s local action on ψ. But acting on ψ surely means changing its phase, so if there is no fact of the matter as to whether the phase of ψ changes then it seems that there is no fact of the matter as to whether **A** acts on ψ. I will discuss this puzzling question in the next three subsections, arguing that we need to understand the interaction between **A** and ψ without invoking gauge-variant notions such as phase changes along open paths.

6.4.1 Healey's Objection

A version of the above objection is found in Healey (1997). It has a peculiar history. When Healey first formulates the argument he seems to advance it against a realist interpretation of the vector potential. In a response, Maudlin (1998) heavily criticises the argument, stating that it "establishes nothing at all". And in his response to Maudlin, Healey suddenly disavows the argument, writing that he "intended not to endorse [it]" (Healey 1999). Although I agree with Maudlin that the argument ultimately fails, I believe that the objection is stronger than both Healey and Maudlin seem to admit. I will therefore elaborate upon it in the remainder of this section.

Here is a reconstruction of the argument:

(i) The vector potential causes a phase difference only if its values change when the solenoid is turned on;

(ii) For any region r which does not enclose the solenoid, there is a gauge such that **A** in r does not change when the solenoid is turned on;

(iii) Therefore, for any region r, there is a gauge such that the potential field in r does not cause a phase difference;

(iv) Leibniz Equivalence: gauge-related models are physically equivalent.

(v) Therefore, for any region r, the potential field in r does not cause a phase difference;

(vi) Therefore, the potential field *anywhere* does not cause a phase difference;

(vii) Therefore, the potential field does not cause a phase difference.

The argument sounds convincing, but Maudlin is correct when arguing that the argument is flawed. It is true that, for any r, there is *a* gauge in which the field at r remains constant when we turn on the solenoid. But of course it is also the case there is *no* gauge for which the field remains constant at *any* r when we turn on the solenoid, since the integral of $\mathbf{A}$ around the solenoid is a gauge-invariant. Therefore, the step from (v) to (vi) is invalid.

But Maudlin's response leaves the exact nature of the interaction between $\mathbf{A}$ and ψ unclear. What is evident from Healey's objection is that it is not invariant whether the phase of ψ changes along an open path. It follows from the Meaningfulness Criterion that statements about the change of phase along open paths are meaningless. On the other hand, the phase different around closed loops *is* invariant and hence meaningful. In other words, according to conventionalism an interaction takes place *somewhere* around the loop, but there are no facts about *where* around the loop the interaction takes place.

I will argue in §6.4.3 that this puzzle rests on a misunderstanding of the concept of interaction in the context of gauge quantities. But first I will give an example which will help us to intuitively understand how we can have a local interaction which nevertheless does not take place at a determinate location.

6.4.2 Intuition Pump/Money Pump

Maldacena (2014) present a thought-provoking analogy between gauge theories and currency exchange. Consider a number of countries, each with their own currency. Some countries have borders between them and at the borders there are exchange banks (see Fig. 6.3). The rule is that when you move across a border you are required to exchange all your bank notes into the currency of the country of arrival.

This set-up has the following local symmetry. Suppose that the exchange rate between two countries is 1 dollar = 10 pesos. We can then decide to 'rescale' the value of the dollar such that one 'old' dollar is now worth two 'new' dollars. Simultaneously, we must change the exchange rate with pesos from 1:10 to 1:5. This generates a new representation which is physically equivalent, as we will see below.

Now suppose that (unlike in the real world) we can earn money by traveling around countries in a loop. For example, suppose that the currencies of the US, Argentina and Italy have the following set of exchange rates:

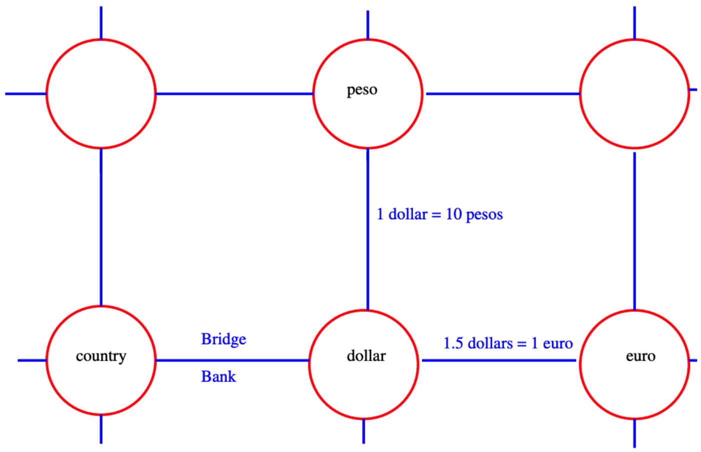

Figure 6.3: Reproduced from Maldacena (2014).

$$1 \text{ dollar} = 1 \text{ peso} \quad 2 \text{ euro} = 1 \text{ peso} \quad 1 \text{ euro} = 1 \text{ dollar} \tag{6.1}$$

In addition, suppose that there are direct borders between all three countries. We can then make money as follows: start off with one dollar in the US, then travel to Argentina and exchange the dollar for one peso. Next, travel to Italy and receive two euros, and finally travel back to the US and exchange these euros for two dollars. We now have double the amount of dollars we started out with!

But *where* did we get rich? At which point during the journey did it become true that we had more money than when we first left home? This question has no meaningful answer. It may seem as if the leg from Argentina to Italy was crucial, since there we exchanged *one* peso for *two* dollars. But recall the gauge symmetry of this set-up. We can rescale the value of the peso such that one peso is now worth ten pesos. The exchange rates then become:

$$1 \text{ dollar} = 10 \text{ pesos} \quad 1 \text{ euro} = 5 \text{ pesos} \quad 1 \text{ euro} = 1 \text{ dollar} \tag{6.2}$$

This gauge transformation does not make a meaningful difference: going round a loop still doubles the amount of dollars we have. It is merely a change of labels. Nevertheless, it is now false that 'nothing' happens when moving from Argentina to Italy. Therefore, it does not matter *where* the number on our bank notes becomes larger of smaller; all that matters it that the *combination* of exchange rates is such that there are non-trivial loops.

There are clear analogies between this scenario and the Aharonov-Bohm effect. In both cases, traveling round a loop results in a measurable difference. And in both cases there is no fact of the matter as to *where* this difference comes about. It is nevertheless clear that the increase in our net wealth is the result of local currency transactions at the borders between countries. There is no need to postulate a holistic causal process which acts along the entire loop 'at once'. Indeed, such an explanation seems absurd. The same is true of the Aharonov-Bohm effect.

6.4.3 Where the Interaction Is

It is now time to address the issue of local action on the conventionalist account directly. Recall the first premise of Healey's argument: (i) the vector potential causes a phase difference only if its value changes when the solenoid is turned on. This sounds like a plausible criterion, but it does not make sense to the conventionalist. After all, the conventionalist denies that it is possible to compare the values of $\mathbf{A}$ across spacetime points. Therefore, there is no fact of the matter about whether the value of $\mathbf{A}$ on a particular open path changes when we turn on the solenoid. Similarly, there are no facts about whether the phase of ψ changes as the electron moves across space.

This does not mean that there is no interaction between them. After all, the integral of $\mathbf{A}$ around the loop enclosing the solenoid *does* change when we turn on the current, as Maudlin points out, and hence there is a measurable phase difference around this loop. The situation here is related to our discussion of the relational nature of the world in Chapter 2. Recall Rovelli's spaceships: if we attempt to describe each fleet on its own, we lose information about the relative distance between them. The world itself is separable, since each ship has a definite spatial location, but our *descriptions* are non-separable, since we have to characterise the world in relative terms. And likewise for the Aharonov-Bohm effect. We cannot describe each open path r in intrinsic terms and then glue our descriptions together, since this leads to the false conclusion that there is no change at all when we turn on the solenoid.[5] Again, we have to describe the intrinsic field values at points *in terms* of their relations to other field values. In the case of electrodynamics, these relations are the (second-order) loop holonomies.

[5]See also Gomes (2019) on the crucial importance of boundaries in gauge theories.

We are thus asking the wrong question when we ask how the local values of **A** and ψ change over time, since these questions have no meaningful answer. Instead, we need to look at the interaction *between* the vector potential field and the wavefunction.

On a literalist account, we can picture this interaction as follows. Consider a section of the fibre bundle of **A**, for example the top one in Fig. 6.2. We can interpret the points of this loop as representing the slope of the phase space of ψ: the highest points of [**A**] represent the steepest slopes. When the electron traces out a circular path, the rate at which its phase changes at each point depends on the slope determined by **A**. The greater the slope at a point, the faster the phase changes. Once the electron has completed its path, the change in its phase is proportional to the holonomy of the field values over that path. In short, the **A**-field acts on ψ by determining how the latter moves through phase space.

The conventionalist account takes this picture as its starting point, but removes the gauge-variant structure. As we have seen, there are no facts about the values of **A** over an open path, and hence no facts about the slope of phase space over such paths. There *are* facts about the total height difference around a closed loop, and these are the holonomies which determine how much ψ changes in the Aharonov-Bohm effect. But the fact that the conventionalist account has less structure does not mean that we cannot understand it as positing a genuine interaction between **A** and ψ. Conventionalism *agrees* with literalism that the values of **A** change the geometrical structure of phase space, but disagrees over *which* changes are physically meaningful. Nevertheless, on both accounts it is the local geometric structure of phase space which determines what happens to the phase of the electron as it moves along its path.

In conclusion, the vector potential field determines the geometrical structure of phase space. As an electron moves around the solenoid, its phase follows a path in phase space; *which* path it follows depends on **A**. But we cannot generate a different path by changing the field values locally, since these changes do not result in qualitative differences: we would merely obtain the *same* path under a different description. From this it follows that there is no fact of the matter about the slope of phase space along open paths, and hence no fact of the matter as to *where* the phase of the electron changes (and by how much). But this does not preclude us from understanding the Aharonov-Bohm effect as the result of ψ's local movement through phase space under the influence of **A**. The world has less structure than we imagined, but stilll enough structure to move around.

6.5 Gauge Realism

Let us return once more to the Inconsistent Triad:

(1) If a quantity is explanatory, it is physically real;

(2) No gauge quantities are physically real;

(3) Some gauge quantities are explanatory.

The orthodox view is that (1) and (2) are true, and hence (3) is false. In this thesis I have given a sustained defence of (3). In Chapter 2 and 3, I made the case against the Invariance Principle, arguing that we need not accept it if we are Quantity-Value Conventionalists. In Chapter 4, I argued that non-invariant quantities often increase the explanatory strength of our theories.

I then turned to *gauge* quantities. I have argued in this chapter that a conventionalist account of gauge quantities offers the best explanation of the Aharonov-Bohm effect. The conventionalist can give a local and deterministic account of the Aharonov-Bohm effect in terms of the vector potential field which explains the cosmic coincidence of holonomy relations. It is both better than the literalist understanding of the vector potential field, and better than a relational holonomy interpretation.

The question I set out to investigate was: *can gauge-dependent quantities play an indispensable role in scientific explanations?* The answer to this question is 'Yes': the vector potential field is indispensable in the best explanation of the Aharonov-Bohm effect. It follows from (1) that the vector potential represents is a physically real field. This settles the Inconsistent Triad in favour of (3).

Chapter 7

Conclusions: a New View on Symmetries

The orthodox view in the philosophy of physics is that symmetries encode superfluous structure. This does not mean that symmetries are insignificant: on this view symmetries are an important guide in interpreting theories. The orthodox view embraces both Leibniz Equivalence, which states that symmetry-related models are physically equivalent, and the Invariance Principle, according to which only invariant quantities are physically real. There are still debates over how these principles are to be applied. For example, the so-called motivationalist (Møller-Nielsen 2017) argues that we cannot invariably interpret non-invariant quantities as redundant structure; we are only allowed to do so once we have found a reduced theory in which these quantities do not occur. These are important questions which promote a critical view of the orthodox position.

But I have argued for an alternative view on symmetries. The orthodox view faces a puzzle: if symmetries are, in a sense, redundant, then why are our best theories gauge theories? This remains an awkward tension of the view that symmetries represent superfluous structure.

This alternative view starts with looking at the aim of interpretation. We interpret theories in order to better understand what the world is like. This is not a statement of scientific realism. For example, a Quine-inspired empiricist (such as I am) believes that the fact that the posit of electrons helps us to organise our experiences is evidence for the existence of electrons. And even the constructive empiricist is interested in the virtues of unobservables, although for her this takes on a more pragmatic dimension. Therefore, even a broad realist commitment implies that explanatory concerns ought to guide our interpretative practises.

But the orthodox view of symmetries ignores the importance of explanation. The Invariance Principles states that non-invariant quantities are unreal, whether or not they are explanatory posits. It is not the case that philosophers have carefully investigated the explanatory advantages of non-invariant quantities but found that there are none. Often, such quantities are dismissed without further consideration. Of course, this is not without reason: on the usual literalist understanding of quantities, interpreting non-invariant quantities as physically real implies theoretical vices such as indeterminism and undetectability. But even then it is important to at least compare these costs with the possible explanatory benefits of interpreting non-invariant quantities as physically real, given the fact that explanation is one of the main reasons for why we interpret theories.

I hope to have offered a way out of this dilemma. If we accept Quantity-Value Conventionalism we can accept Leibniz Equivalence without committing to the Invariance Principle. This allows us to identify symmetry-related models and thereby avoid indeterminism, while at the same time maintaining a realist interpretation of non-invariant quantities. This position is an extension of sophisticated substantivalism to non-spatiotemporal properties, and it has a perspicuous metaphysical picture in the form of anti-quidditism.

I first illustrated Quantity-Value Conventionalism with the debate between the absolutist and the comparativist about mass, but the real import of this position lies in its application to gauge theories. In the final two chapters of this thesis I have given a detailed case study of the Aharonov-Bohm effect in which I argued that the choice between absolutism and relationism about gauge quantities is a false dilemma. It is well-known that the former option leads to indeterminism, and I have argued that the latter implies a cosmic coincidence which violates local action. But the orthodox Invariance Principle implies that we either have to consider gauge-related models as physically distinct, or embrace a relational ontology. Quantity-Value Conventionalism offers a third option. I have argued that this unexplored position in the space of possibilities may well represent the best of both worlds.

This new view on symmetries thus satisfies a demand for explanation which the orthodoxy neglects. It also offers a new perspective on the role of symmetries in our theories. According to the conventionalist approach, the fact that our theories admit of physically equivalent models is a *consequence* of the fact that quantities have no intrinsic identities. If this is the case we can only describe theories in relational terms, and so it follows that transformations which leave these relations invariant generate physically equivalent models. There is thus a sense in which theories with invariance symmetries *do* contain redundant structure, since their representational relation to the world is many-to-one. But these redundancies are not a theoretical vice. On the contrary, the fact that our representations of nature are redundant tells us something interesting about its fundamental structure. The conventionalist approach explains *why* our best theories have symmetries: because the fundamental quantities of the world lack intrinsic identities. Furthermore, our best theories are *gauge* symmetries because the fundamental quantities are incomparable across spacetime points.

In closing, I will consider the tasks ahead. First, there is work to be done in spelling out the finer details of Quantity-Value Conventionalism. The application of this approach to gauge quantities is a novel and versatile suggestion which deserves further exploration.

There is also much room for further application of the conventionalist approach. I have given a detailed account of the vector potential field, but there are other gauge quantities whose status is disputed. For example, it has recently been debated whether gravitational energy in general relativity is physically significant despite its gauge-variance (see Read (2018) and Duerr (2018)). Quantity-Value Conventionalism may enable a realist interpretation of gravitational energy which preserves its explanatory power. We can also use conventionalist to explain relativistic effect such as time dilation and length contraction in terms of non-invariant forces. This is reminiscent of the so-called dynamical approach to spacetime theories, which is a further connection to explore. Finally, we have not yet looked at Quantity-Value Conventionalism as applied to quantum theories, but of course there too symmetries play an important role. There is thus much scope for further research once we recognise conventionalism as an original interpretative option.

Bibliography

Aharonov, Y. and D. Bohm. 1959. Significance of Electromagnetic Potentials in the Quantum Theory. *Physical Review 115*(3): 485–491.

Arntzenius, F. 2012. *Space, Time, and Stuff.* Oxford, New York: Oxford University Press.

Asselmeyer-Maluga, T. 2016. Smooth quantum gravity: Exotic smoothness and Quantum gravity. *arXiv:1601.06436 [gr-qc, physics:hep-th, physics:math-ph]*.

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

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

Belot, G. 2003. Notes on Symmetries. In *Symmetries in Physics: Philosophical Reflections*, eds. K. A. Brading and E. Castellani, 393–412. Cambridge University Press.

Belot, G. 2013. Symmetry and Equivalence. In *The Oxford Handbook of Philosophy of Physics*, ed. R. Batterman, 318–339. Oxford University Press.

Bostrom, N. 2003. Are We Living in a Computer Simulation? *The Philosophical Quarterly 53*(211): 243–255.

Bricker, P. 2017. Is There a Humean Account of Quantities? *Philosophical Issues 27*(1): 26–51.

Brown, H. 2016. The (magnetic) Aharonov-Bohm effect [lecture notes].

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.

Dasgupta, S. 2013. *Absolutism vs Comparativism about Quantity.* Oxford University Press.

Dasgupta, S. 2016. Symmetry as an Epistemic Notion (Twice Over). *The British Journal for the Philosophy of Science 67*(3): 837–878.

Dees, M. 2015. Maudlin on the Triangle Inequality. *Thought: A Journal of Philosophy 4*(2): 124–130.

Dees, M. 2018. Physical Magnitudes. *Pacific Philosophical Quarterly 99*(4): 817–841.

Denby, D. A. 2001. Determinable Nominalism. *Philosophical Studies 102*(3): 297–327.

Dewar, N. 2015. Symmetries and the Philosophy of Language. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics 52*(Part B): 317–327.

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

Dirac, P. A. M. 1930. *The Principles of Quantum Mechanics*. Clarendon Press.

Duerr, P. M. 2018. Fantastic Beasts and Where (Not) to Find Them: Local Gravitational Energy and Energy Conservation in General Relativity. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*.

Earman, J. 2017. The role of idealizations in the Aharonov–Bohm effect. *Synthese*: 1–29.

Earman, J. and J. Norton. 1987. What Price Spacetime Substantivalism? The Hole Story. *The British Journal for the Philosophy of Science 38*(4): 515–525.

Eddon, M. 2013. Quantitative Properties. *Philosophy Compass 8*(7): 633–645.

Ellis, B. 1966. *Basic Concepts of Measurement*. Cambridge University Press.

Feynman, R. P., R. B. Leighton, and M. Sands. 1964. *The Feynman Lectures on Physics*, Volume II. Reading: Addison-Wesley.

Funkhouser, E. 2006. The Determinable-Determinate Relation. *Noûs 40*(3): 548–569.

Gomes, H. 2019. Gauging the Boundary in Field-space. *arXiv:1902.09258 [hep-th, physics:physics]*.

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

Hacking, I. 1981. Do We See Through a Microscope? *Pacific Philosophical Quarterly 62*(4): 305–322.

Harker, D. 2010. Two Arguments for Scientific Realism Unified. *Studies in History and Philosophy of Science Part A 41*(2): 192–202.

Healey, R. 1997. Nonlocality and the Aharonov-Bohm Effect. *Philosophy of Science 64*(1): 18–41.

Healey, R. 1999. Quantum Analogies: A Reply to Maudlin. *Philosophy of Science 66*(3): 440–447.

Healey, R. 2006. Symmetry and the Scope of Scientific Realism. In *Physical Theory and Its Interpretation: Essays in Honor of Jeffrey Bub*, eds. W. Demopoulos and I. Pitowsky, The Western Ontario Series in Philosophy of Science, 143–160. Dordrecht: Springer Netherlands.

Healey, R. 2007. *Gauging What's Real: The Conceptual Foundations of Contemporary Gauge Theories*. Oxford, New York: Oxford University Press.

Healey, R. A. 1994. Nonseparable processes and causal explanation. *Studies in History and Philosophy of Science 25*(3): 337–374.

Hesse, M. B. 1955. Action at a Distance in Classical Physics. *Isis 46*(4): 337–353.

Ismael, J. and B. C. Van Fraassen. 2003. Symmetry as a Guide to Superfluous Theoretical Structure. In *Symmetries in Physics: Philosophical Reflections*, eds. K. Brading and E. Castellani, 371–92. Cambridge University Press.

Kitcher, P. 1993. *The Advancement of Science: Science Without Legend, Objectivity Without Illusions*. Oxford University Press.

Kosso, P. 2000. The Empirical Status of Symmetries in Physics. *The British Journal for the Philosophy of Science 51*(1): 81–98.

Laudan, L. 1981. A Confutation of Convergent Realism. *Philosophy of Science 48*(1): 19–49.

Leeds, S. 1999. Gauges: Aharonov, Bohm, Yang, Healey. *Philosophy of Science 66*(4): 606–627.

Lewis, D. 1970. How to Define Theoretical Terms. *Journal of Philosophy 67*(13): 427–446.

Lewis, D. 1973. Causation. *Journal of Philosophy 70*(17): 556–567.

Lewis, D. 1979. Counterfactual Dependence and Time's Arrow. *Noûs 13*(4): 455–476.

Lyre, H. 2004. Holism and Structuralism in U(1) Gauge Theory. http://philsci-archive.pitt.edu/1831/.

Maldacena, J. 2014. The symmetry and simplicity of the laws of physics and the Higgs boson. *arXiv:1410.6753 [hep-ex, physics:hep-lat, physics:hep-ph, physics:hep-th, physics:physics]*.

Martens, N. and J. Read. Forthcoming. Sophistry About Symmetries?

Martens, N. C. M. 2019a. Machian Comparativism about Mass. *The British Journal for the Philosophy of Science*.

Martens, N. C. M. 2019b. The (Un)detectability of Absolute Newtonian Masses. http://philsci-archive.pitt.edu/15919/.

Mattingly, J. 2006. Which Gauge Matters? *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics 37*(2): 243–262.

Maudlin, T. 1998. Healey on the Aharonov-Bohm Effect. *Philosophy of Science 65*(2): 361–368.

Maudlin, T. 2007. Suggestions from Physics for Deep Metaphysics. In *The Metaphysics Within Physics*. Oxford University Press.

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

Møller-Nielsen, T. and J. Read. Forthcoming. Redundant Epistemic Symmetries.

Mundy, B. 1986. On the General Theory of Meaningful Representation. *Synthese 67*(3): 391–437.

Mundy, B. 1987. The Metaphysics of Quantity. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition 51*(1): 29–54.

Nounou, A. M. 2003. A Fourth Way to the Aharonov-Bohm Effect. In *Symmetries In Physics: Philosophical Reflections*, eds. K. Brading and E. Castellani. Cambridge University Press.

Nounou, A. M. 2010. Holonomy Interpretation and Time: An Incompatible Match? A Critical Discussion of R. Healey's Gauging What's Real: The Conceptual Foundations of Contemporary Gauge Theories. *Erkenntnis (1975-) 72*(3): 387–409.

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

Peshkin, M. and A. Tonomura. 1989. *The Aharonov-Bohm Effect*. Berlin: Springer.

Pooley, O. 2012. Substantivalist and Relationalist Approaches to Spacetime. http://philsci-archive.pitt.edu/9055/.

Psillos, S. 1999. *Scientific Realism: How Science Tracks Truth*. Routledge.

Putnam, H. 1975. *Mathematics, Matter and Method: Volume 1, Philosophical Papers*. CUP Archive.

Quine, W. V. 1966. Posits and Reality. In *The Ways of Paradox*. New York: Random.

Read, J. 2018. Functional Gravitational Energy. *The British Journal for the Philosophy of Science*.

Read, J. and T. Møller-Nielsen. 2018. Motivating Dualities. *Synthese*.

Rovelli, C. 2014. Why Gauge? *Foundations of Physics 44*(1): 91–104.

Ruetsche, L. 2011. *Interpreting Quantum Theories*. Oxford, New York: Oxford University Press.

Saatsi, J. 2005. Reconsidering the Fresnel–Maxwell Theory Shift: How the Realist Can Have Her Cake and EAT It Too. *Studies in History and Philosophy of Science Part A 36*(3): 509–538.

Saunders, S. 2003. Physics and Leibniz's Principles. In *Symmetries in Physics: Philosophical Reflections*, eds. K. Brading and E. Castellani, 289–307. Cambridge University Press.

Saunders, S. 2007. Mirroring as an A Priori Symmetry. *Philosophy of Science 74*(4): 452–480.

Skow, B. 2011. Does Temperature Have a Metric Structure? *Philosophy of Science 78*(3): 472–489.

Smart, J. J. C. 1963. *Philosophy And Scientific Realism*. Humanities Press.

Stalnaker, R. 1979. Anti-Essentialism. *Midwest Studies In Philosophy 4*(1): 343–355.

Suppes, P. 1960. A Comparison of the Meaning and Uses of Models in Mathematics and the Empirical Sciences. *Synthese 12*(2-3): 287–301.

Trautman, A. 1980. Fiber Bundles Gauge Fields and Gravitation. In *General Relativity and Gravitation. Vol. 1. One Hundred Years after the Birth of Albert Einstein. Edited by A. Held. New York, NY: Plenum Press, p.287, 1980*, 287.

Van Fraassen, B. C. 1967. Meaning Relations among Predicates. *Noûs 1*(2): 161–179.

Van Fraassen, B. C. 2008. *Scientific Representation: Paradoxes of Perspective*. Oxford, New York: Oxford University Press.

Wallace, D. 2002. Time-Dependent Symmetries: The Link Between Gauge Symmetries and Indeterminism. In *Symmetries in Physics: Philosophical Reflections*, eds. K. Brading and E. Castellani, 163–173. Cambridge University Press.

Wallace, D. 2014. Deflating the Aharonov-Bohm Effect. *arXiv:1407.5073 [quant-ph]*.

Wallace, D. 2015. Fields as Bodies: A unified presentation of spacetime and internal gauge symmetry. *arXiv:1502.06539 [gr-qc, physics:hep-th]*.

Weatherall, J. O. 2011. On (Some) Explanations in Physics. *arXiv:1106.2334 [physics]*.

Weatherall, J. O. 2014. Fiber Bundles, Yang-Mills Theory, and General Relativity. *arXiv:1411.3281 [gr-qc, physics:hep-th, physics:math-ph, physics:physics]*.

Weatherall, J. O. 2015. Understanding Gauge. *arXiv:1505.02229 [physics]*.

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